

Unit 8

Teaching how to read

Introduction

For most of the data you have met in previous units, the numbers in the dataset have had an intrinsic meaning. For example, the miles per gallon that a particular car achieved (Unit 1, Section 3), the price of a jar of coffee (Unit 2, Section 1) or the gross hourly earnings of men and women (Unit 3, Section 1). Some of the statistical techniques you have already learned about, such as the z -tests in Unit 7, are appropriate for such data. In contrast, Unit 8 focuses on data where numbers are used only as labels (or not used at all) – and on statistical methods specifically designed for such data.

The theme of this unit is learning how to read and, more particularly, which teaching methods produce the best results. Section 1 provides some background, clarifies the questions of interest, and describes a large study to investigate different methods of teaching how to read. The data from the study are in *categorical form*, where the categories are determined, for example, by the teaching method used, or whether a given standard has been met or not. The data are presented in tables of a particular type, called *contingency tables*.

Section 2 describes how contingency tables are constructed, and how they are used to study relationships between variables. In Section 3, you will learn more about probabilities of various kinds, and how they are related. Then in Section 4 a hypothesis test known as the χ^2 *test for contingency tables* is introduced. This follows a similar procedure to that in the previous two units. We set up null and alternative hypotheses and then calculate a test statistic. We then compare the test statistic with critical values at the 5% and 1% significance levels and so decide whether or not to reject the null hypothesis. Section 5 makes further use of this test, with particular reference to the ideas of association and causality in assessing the impact of teaching methods on literacy; it also discusses some reservations about the conclusions from a large study investigating methods for teaching how to read, and about conclusions from hypothesis tests in general. Finally, Section 6 directs you to the Computer Book which describes how to use Minitab to analyse data in the form of contingency tables.

The symbol χ is the Greek letter chi, pronounced *ki* (to rhyme with tie); χ^2 is pronounced *ki squared*.

1 The best way of teaching how to read

Reading is a key skill that children are expected to master early in their schooling. So an important question for educators, and the question that forms the theme for this unit, is:

What is the best way of teaching how to read?

In this unit, you will mainly consider data from one study designed to help answer this question. In Subsection 1.2, the question will be clarified, and in Subsection 1.3 the study itself will be described. However, first Subsection 1.1 provides some background.



1.1 The 'reading wars'

Unlike languages such as Spanish, Finnish and Swahili, where there is a direct correspondence between how words sound and how they are written, English spelling is far from transparent – for example, the 'o' is pronounced differently in 'grove', 'move' and 'top', while the spelling of words like 'bomb', 'knot', 'enough' and 'aisle' defies logic. This feature of the English language has generated much debate about how best to teach reading at primary school and in particular what emphasis to place on relating the letters of the alphabet to the individual sounds that words are made up of.

Phonemes and graphemes

A **phoneme** is a term used to describe an individual sound that can distinguish one word from another.

A **grapheme** is a term used to describe a letter, or combination of letters, of the alphabet corresponding to an individual sound.

For example, the word ‘cat’ comprises three phonemes corresponding to the sounds associated with the three graphemes (individual letters in this case) ‘c’, ‘a’ and ‘t’.

Choosing which teaching method to use has been very controversial, giving rise to what has been called the ‘reading wars’. Several reviews have established that it is important for children to be able to use grapheme–phoneme, or letter–sound, correspondences, also known as **phonics**. Thus teaching methods based solely on learning whole words have largely been superseded by methods that involve teaching phonics.

However, there are several teaching approaches that involve phonics, and so the debate continues. One is a ‘top-down’ approach in which children are taught to recognise whole words before *analysing* how they can be broken down into grapheme–phoneme correspondences. This approach, which is called **analytic phonics**, was widely used in Britain until relatively recently. A contrasting approach is to start with intensive teaching of how letters of the alphabet relate to phonemes, and building these up – *synthesizing* them – into whole words. This approach is called **synthetic phonics**, and has recently been introduced in many schools across Britain.

Here is Peter, here is Jane

If you learned how to read in Britain in the 1960s, the chances are that you will have met Peter and Jane, the main characters of the books published by Ladybird as part of the Key Words Reading Scheme (Figure 1). This teaching method, also known as ‘Look and Say’, was based entirely on learning whole words. Using a limited vocabulary with much repetition – resulting in stilted prose and very odd-sounding conversations – the aim was to teach children to recognise the most commonly used words.



Figure 1 The first of the Peter and Jane series, published in 1964

1.2 Clarifying the question

Subsection 1.3 will describe a large study that was conducted in order to decide which method is most effective for teaching how to read. We will keep returning to this study throughout the unit. However, the question *What is the best way of teaching how to read?* is a very general one, and so needs to be clarified before it can be addressed.

Activity 1 Clarifying a question

Briefly think about the question *What is the best way of teaching how to read?*, and write down two or three of its key aspects which *you* think need to be clarified. (The wording of the question is a useful guide.)

1. Clarify

In Subsection 1.1, two contrasting methods for teaching how to read at school were described: analytic phonics and synthetic phonics. Thus it would seem natural to design a study to compare these two distinct approaches.

Reading ability comprises many facets, which cannot easily be reduced to a single measurement: thus several aspects of reading ability ought to be evaluated. One standardised single-word reading test, called the British Ability Scales Word Reading Test, or BAS test for short, provides a standardised reading score. This test enables the reading ability of a child aged under 14.5 years old to be classified according to whether or not the child's reading age is above their chronological age (so the test tells you whether or not a child is reading well for their age). Other tests may be used to assess reading comprehension, spelling ability, and other aspects of reading skill.

The effectiveness of a method for teaching how to read can be judged by the proportion of children whose reading age is higher than their chronological age, as measured by the BAS test, for example. The BAS test was also used as a measure of reading ability in Unit 7, where results from the British Cohort Study were examined.

1.3 The Clackmannanshire study

The data in this unit come from an influential study undertaken by Rhona Johnston and Joyce Watson over seven years in several schools in Clackmannanshire (Figure 2) involving some 300 children in 13 classes. (Johnston and Watson (2004) ‘Accelerating the development of reading, spelling and phonemic awareness skills in initial readers’, *Reading and Writing: An Interdisciplinary Journal*, vol. 17, pp. 327–357.) This study compared three methods for teaching how to read in the first year of primary school. Two of the methods were analytic phonics and synthetic phonics, which were described in Subsection 1.1. In addition, a third teaching method was evaluated, which combined analytic phonics with training in phonological awareness (the ability to recognise the different sounds of which spoken words are made up). This method will be referred to as analytic phonics + PA. Eleven different aspects of reading ability were evaluated for each child, including the BAS reading test described in Subsection 1.2.

Clearly, teaching methods are applied to whole classes, rather than to individual children. Some care in selecting the participating classes and schools was needed to ensure that the children within the three different groups (taught using analytic phonics, taught using synthetic phonics and taught using analytic phonics + PA) were broadly comparable and that any imbalances were known about in advance.



Figure 2 Clackmannanshire, at the heart of Scotland

2. Collect

Example 1 Comparability across the groups

All the children within the selected classes were 5-year-old new school entrants, whose first language was English. Socio-economic background may influence reading ability, so information on such factors was obtained. In Clackmannanshire, schools are classified according to an index of disadvantage ranging from -1 (least deprived) to 2.5 (most deprived). The analytic phonics group scored -0.13 on this index, whereas the analytic phonics + PA group scored 0.05 , and the synthetic phonics group 1.12 . Thus, the children in the synthetic phonics group came from slightly more disadvantaged socio-economic backgrounds than children from the other groups.

Activity 2 Making comparisons at baseline

In Example 1, some differences in socio-economic background between the groups were described. There may well be other factors that might have influenced reading ability even before formal teaching began. Such differences between groups at the start of the study (‘at baseline’) would mean that we were not comparing like with like.

Briefly suggest how to check that, at the start of the study, the children have comparable reading ability.

In the Clackmannanshire study, the children were tested two weeks after starting school, but before the teaching of reading started. Then the three teaching programmes were introduced, lasting 16 school weeks, after which the children were tested again. Further testing took place subsequently, to monitor the children's progress.

Exercise on Section 1

Exercise 1 *Investigating differences between girls and boys*

There is some evidence to suggest that boys and girls learn how to read differently, and may thus respond differently to different teaching methods. Briefly outline how such effects might be investigated.



"It's called 'reading'. It's how people
install new software into their brains"

2 Contingency tables

In the Clackmannanshire study described in Subsection 1.3, each child was classified according to whether their reading age (as measured by the BAS reading test) was higher than their chronological age (that is, ‘better than expected for their age’), or whether it was less than or equal to their chronological age. This gives rise to a table of counts called a *contingency table*.

2.1 Tables and proportions

Example 2 *The baseline test results*

The results of the baseline test in the Clackmannanshire study (prior to starting the teaching programme) are shown in Table 1.

Table 1 Results from the baseline test

	Reading age as compared to chronological age	
	Not higher	Higher
Analytic phonics	69	39
Analytic phonics + PA	43	35
Synthetic phonics	76	41

The three rows of numbers in the table correspond to the three groups, and the two columns of numbers correspond to reading ability, classified as reading age less than or equal to, or higher than, chronological age. The numbers in the table are numbers of children. So, for example, 69 children in the analytic phonics group had a reading age which was equal to or less than their chronological age.

One question of interest, for the baseline test data, is to check that any differences between the groups are small. (We would expect to see some differences, owing to random fluctuations.) You learned a way to quantify such differences in Unit 6 using probabilities; this is the topic of Example 3.

3. Analyse

Example 3 *Calculating probabilities from the baseline test results*

To calculate probabilities, it's convenient to have the totals available, for both the rows and the columns. These totals, which are called **marginal totals**, are shown in Table 2 (which will be referred to at various points in the unit).

Table 2 Results from the baseline test, with marginal totals

	Reading age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	69	39	108
Analytic phonics + PA	43	35	78
Synthetic phonics	76	41	117
Total	188	115	303

We obtain probabilities from a contingency table by calculating *proportions*. For example, if a child in the analytic phonics (AP) group is selected at random, then the probability that a child's reading age is higher than their chronological age is:

$$\begin{aligned} & \frac{\text{number in AP group with reading age} > \text{chronological age}}{\text{number in AP group}} \\ &= \frac{39}{108} \simeq 0.361. \end{aligned}$$

Multiplying this result by 100 gives about 36.1%.

Also of interest is the overall proportion of children who could read well for their age, before beginning to learn how to read at school. This can be calculated using the column totals. So the probability a child in the study has reading age above their chronological age is:

$$\begin{aligned} & \frac{\text{total number with reading age} > \text{chronological age}}{\text{total number in study}} \\ &= \frac{115}{303} \simeq 0.380. \end{aligned}$$

Multiplying this result by 100 gives about 38.0%.



Activity 3 *Comparing proportions at baseline*

Use the data in Table 2 to do the following.

- (a) Obtain the percentage of children in the analytic phonics + PA group whose reading age is higher than their chronological age. Do the same for children in the synthetic phonics group.
- (b) Compare these proportions informally. What do you conclude about the comparability of the three groups?

In Activity 3 you found some moderate differences between groups in the baseline test; this is not unexpected, as some differences will arise from random fluctuations. The question is whether the differences that were observed could plausibly have arisen due to chance. This topic will be addressed in Section 4.

As previously mentioned, the children were tested again after 16 school weeks of teaching. The results are given in Example 4, and discussed in Activity 4.

Example 4 *The first follow-up test results*

Table 3 shows the contingency table of results on the BAS reading test at the first follow-up test for the Clackmannanshire study after 16 school weeks. (Table 3 will be referred to at various points in the unit.)

Table 3 Results from the first follow-up test

	Reading age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	68	36	104
Analytic phonics + PA	51	24	75
Synthetic phonics	35	78	113
Total	154	138	292

Note that the total numbers in each group have dropped a little, perhaps because some children changed schools or moved out of the area. The biggest change, however, is in the numbers of children performing well in the synthetic phonics group.



Learning how to read

Activity 4 *Comparing proportions at first follow-up*

Use the data in Table 3 to do the following.

- (a) Obtain the percentages of children in the three groups whose reading age at the first follow-up test is greater than their chronological age.
- (b) Compare these percentages informally.
- (c) What further information might you require to interpret these results?



Activity 4 suggests that, for reading ability as measured by the BAS reading test, synthetic phonics is a more effective teaching method than analytic phonics or analytic phonics supplemented with phonological awareness training. However, as indicated after Activity 3, a more formal hypothesis test is required to interpret such data.

Contingency tables are an effective way of displaying such data, and statistical methods have been developed to analyse them. You will encounter some of these methods in this unit. (Note that the published

analysis of the Clackmannanshire study used more complex methods. However, the conclusions are the same as those that will be made in this unit.)

2.2 Identifying contingency tables

We have now introduced some of the data, in the form of contingency tables, to help us investigate the question clarified in Subsection 1.2: *What is the best way of teaching how to read?* In this subsection we shall look more closely at the form of the data presented in Tables 2 and 3 in Subsection 2.1. For ease of reference, Table 3 is reproduced here.

Table 4 Results from the first follow-up test (reproduced from Table 3)

	Reading age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	68	36	104
Analytic phonics + PA	51	24	75
Synthetic phonics	35	78	113
Total	154	138	292

In this table, each child in the sample of 292 children is assigned to one of three teaching method categories, and to one of two reading age categories. The teaching method and reading age variables are described as **categorical** variables, as they are each made up of a certain number of categories; each child can be classified into one of these categories. In particular, we are not told the children’s actual scores on the BAS reading test, or their reading ages, so the data are of a different form from the data that we used when we applied the *z*-test in Unit 7.

In Table 4, the categories for each of the variables are mutually exclusive. (You met this expression in Unit 6 (Subsection 2.2) when discussing probabilities of events and the addition rule.) This means that each child can belong to one and only one category on each variable.

Contingency tables

A contingency table is a table which meets the following three conditions:

- the row variable and the column variable are both *categorical*
- the categories for both variables are *mutually exclusive*
- the entry in each cell of the table is a *count*.

Example 5 *Contingency table or not?*

In Activity 4 (Subsection 2.1) you obtained the percentages of children in each teaching group whose reading age at the first follow-up test was higher than their chronological age. These percentages, together with those of children with reading age not higher than chronological age, are shown in Table 5.

Table 5 First follow-up test: percentages with reading age higher/not higher than chronological age

	Reading age as compared to chronological age	
	Not higher (%)	Higher (%)
Analytic phonics	65.4	34.6
Analytic phonics + PA	68.0	32.0
Synthetic phonics	31.0	69.0

The variables displayed in Table 5 are categorical, and the categories are mutually exclusive. However, the entries are percentages, not counts. Hence this table is *not* a contingency table.

Table 6 provides a little more insight into how many children progressed between tests. (In this table we have replaced ‘Not higher’ by ‘Low’ and ‘Higher’ by ‘High’ to shorten the column headings.)

Table 6 Reading ability at baseline and at first follow-up test

	Reading ages at baseline/follow-up				Total
	Low/Low	Low/High	High/Low	High/High	
Analytic phonics	50	16	17	20	103
Analytic phonics + PA	30	10	21	14	75
Synthetic phonics	23	52	12	26	113
Total	103	78	50	60	291

Thus, for example, in the analytic phonics group, 50 children scored Low on both tests, 16 scored Low on the baseline test and High on the first follow-up test, 17 scored High on the baseline test and Low on the first follow-up test, and 20 scored High on both tests. Note that the baseline values are missing for one child, so the total is 291 rather than 292 as in Table 4.

Is this table a contingency table? The variables are certainly categorical, and the categories are mutually exclusive, since each child can only be in one of the four cells Low/Low, Low/High, High/Low, High/High. And finally, the entries are counts. So yes, this is a contingency table.



Children reading ... or just looking at the pictures?

Recognising contingency tables is important, because the statistical test described in Section 4 applies *only* to contingency tables. Activity 5 will give you some more practice at identifying which tables are contingency tables and which are not.

Activity 5 *Identifying contingency tables*

Say whether the following tables are contingency tables, stating reasons why or why not.

- (a) Table 7 gives the average ages of the children in the three teaching groups at baseline test and at first follow-up test.

Table 7 Average chronological age of children at baseline and first follow-up tests

	Chronological age at test (years)	
	Baseline test	First follow-up test
Analytic phonics	5.01	5.43
Analytic phonics + PA	5.01	5.41
Synthetic phonics	5.00	5.48

- (b) Table 8 shows the numbers of children (among those who did the follow-up test) whose reading age was higher than their chronological age at the baseline and at the first follow-up tests. (Note: this table was derived from Table 6.)

Table 8 Reading ability at baseline and at first follow-up test

	Standardised reading age at baseline and first follow-up	
	High at baseline	High at first follow-up
Analytic phonics	37	36
Analytic phonics + PA	35	24
Synthetic phonics	38	78

A contingency table is usually described in terms of the number of categories of each variable. For example, in Table 4, the *teaching method* variable has three categories, and the *reading ability* variable has two categories. This table is therefore said to be a

3×2 contingency table (where ‘ 3×2 ’ is read as ‘three by two’).

In general, if the variable corresponding to the rows has r categories and the variable corresponding to the columns has c categories, the contingency table is described as an **$r \times c$ contingency table**. It may also be said to be of **size $r \times c$** , or of **dimension $r \times c$** .

Activity 6 *Determining sizes of contingency tables*

- (a) Write down the dimension of Table 6 from Example 5.
- (b) Suppose that rows and columns were swapped around in Table 4. What would be the dimension of the swapped-round version of the table?

Exercises on Section 2

Exercise 2 *Examining a gender distribution*

Table 9 shows the gender distribution by teaching group in the Clackmannanshire study.



Table 9 Gender distribution by teaching group

	Gender		
	Boys	Girls	Total
Analytic phonics	58	51	109
Analytic phonics + PA	39	39	78
Synthetic phonics	61	56	117
Total	158	146	304

- (a) Obtain the percentage of girls in the study as a whole, to one decimal place.
- (b) Obtain the proportions of boys in each teaching group.
- (c) Comment briefly (and informally) on any differences you observe between groups.

Exercise 3 Contingency or not contingency?

Tables 10 and 11 below were both derived from Table 9. In Table 10, the category ‘analytic phonics with or without PA’ is obtained by combining the categories ‘analytic phonics’ and ‘analytic phonics + PA’ from Table 9.

Table 10 Gender distribution by teaching group

	Gender		Total
	Boys	Girls	
Analytic phonics with or without PA	97	90	187
Synthetic phonics	61	56	117
Total	158	146	304

Table 11 Gender distribution by teaching group (%)

	Gender	
	Boys	Girls
Analytic phonics	53	47
Analytic phonics + PA	50	50
Synthetic phonics	52	48

- (a) For each table, say whether it is a contingency table, giving reasons for your answer.
- (b) Write down the dimension of each of the contingency table(s) you identified in part (a).

3 Calculating probabilities

‘To us, probability is the very guide of life’ – attributed to Cicero (106–43 BC).



A bust of Cicero

In this section, you will learn more about calculating probabilities, following on from the material in Unit 6. However, first the next phase of the Clackmannanshire study is described.

Example 6 The second follow-up test

At the first follow-up test, as suggested in Example 4 (Subsection 2.1), it became apparent that children in the synthetic phonics group were progressing much more rapidly than those in the other two teaching groups. For ethical reasons, it would be unacceptable to withhold a more effective teaching programme. Thus, from then on, all children were taught using the synthetic phonics method. At the end of the following school year, all the children were retested once more. The results of this second follow-up test, using the BAS reading score as before, are shown in Table 12; the row variable now describes the teaching group to which the children were originally allocated.

Table 12 Results from the second follow-up test

	Reading age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	19	78	97
Analytic phonics + PA	11	55	66
Synthetic phonics	15	90	105
Total	45	223	268

It appears from Table 12 (which will be referred to at various points in the unit) that the three teaching groups are much more similar at the second follow-up test than they were at the first follow-up test. Thus the proportion of children initially allocated to the analytic phonics group whose reading age is higher than their chronological age is

$$\frac{78}{97} \simeq 0.804,$$

while it is

$$\frac{55}{66} \simeq 0.833$$

for children initially allocated to the analytic phonics + PA group, and

$$\frac{90}{105} \simeq 0.857$$

for children initially allocated to the synthetic phonics group.

Thus the children in the two analytic phonics groups appear to have ‘caught up’ with their peers in the synthetic phonics group, following the decision to include all children in the synthetic phonics programme.

At this point we shall pause the teaching story, and consider in a little more detail different types of probabilities (proportions), including those calculated informally in Example 6.

3.1 Joint and conditional probabilities

In this subsection, we shall make an important distinction between probabilities involving two events A and B .

The **joint probability of A and B** , denoted $P(A \text{ and } B)$, is the probability that both A and B occur together. In contrast, the **conditional probability of A given B** , denoted $P(A|B)$, is the probability that A occurs given that B occurs. Note that we do not require A and B to be mutually exclusive in these definitions.

Example 7 shows how these definitions work for the data from the second follow-up test of the Clackmannanshire study.

Example 7 *Joint and conditional probabilities at the second follow-up*

In order to avoid cumbersome descriptions, let us first label events as follows.

- Event R_2 : reading age higher than chronological age at the second follow-up test.
(Similarly, R_0 and R_1 will refer to reading age being higher than chronological age at the baseline and first follow-up tests, respectively.)
- Event AP : initially allocated to the analytic phonics group.
- Event $AP+$: initially allocated to the analytic phonics + PA group.
- Event SP : initially allocated to the synthetic phonics group.

Suppose we want the probability that a child in the study was in the AP group *and* had a high reading age. That is, we wish to estimate the joint probability of events R_2 and AP for a child chosen randomly from the study. Using the approach described in Unit 6 and applying it to the data from Table 12, this is:

$$\begin{aligned} P(R_2 \text{ and } AP) &= \frac{\text{number of children with event } R_2 \text{ and event } AP}{\text{total number of children}} \\ &= \frac{78}{268} \simeq 0.291, \end{aligned}$$

or about 29.1%.

Now instead suppose we list all the children in the AP group and randomly choose a child from this list. The probability that the chosen child experienced event R_2 is the conditional probability that a randomly chosen child experiences event R_2 , given that the child experiences event AP . More succinctly, it is the conditional probability of event R_2 , given event AP . The only children entering into the calculation are those who have experienced event AP . Hence,

$$\begin{aligned} P(R_2|AP) &= \frac{\text{number of children experiencing } R_2 \text{ and } AP}{\text{number of children experiencing } AP} \\ &= \frac{78}{97} \simeq 0.804, \end{aligned}$$

or about 80.4%.

Note that $P(R_2 \text{ and } AP)$ and $P(R_2|AP)$ have the same numerator, but different denominators: for the conditional probability, the denominator is restricted to children experiencing event AP .

Usually, for any two events A and B , $P(A|B)$ is not equal to $P(B|A)$. To demonstrate this, we calculate the conditional probability $P(AP|R_2)$ when we restrict our attention to those children who have a high reading age, pick one of them at random, and want the probability that the chosen child is from the AP group. So now the only children entering into the calculation are those who have experienced R_2 :

$$\begin{aligned}
 P(AP|R_2) &= \frac{\text{number of children experiencing } AP \text{ and } R_2}{\text{number of children experiencing } R_2} \\
 &= \frac{78}{223} \simeq 0.350,
 \end{aligned}$$

or about 35.0%. This is very different to $P(R_2|AP)$.

Activity 7 will give you some practice at working out joint and conditional probabilities.

Activity 7 *Calculating joint and conditional probabilities*



Use the data in Table 12 (Section 3) to calculate the following probabilities. The event labels are as in Example 7.

- (a) $P(R_2 \text{ and } SP)$.
- (b) $P(R_2|SP)$.
- (c) $P(AP+|R_2)$.

The notation $P(A \text{ and } B)$ and $P(A|B)$ makes it explicit whether we are referring to a joint probability or a conditional probability. When probabilities are described in words, however, care is needed to make sure you have correctly identified what type of probability is involved. This is the topic of Activity 8.

Activity 8 *Identifying joint and conditional probabilities*

The following phrases describe probabilities involving the events R_2 and SP described in Example 7. In each case, write down whether the probability is $P(R_2 \text{ and } SP)$, $P(R_2|SP)$, or $P(SP|R_2)$.

- (a) The probability that a randomly selected child who was initially allocated to the synthetic phonics group has a reading age higher than their chronological age at the second test.
- (b) The probability that a randomly selected child was initially allocated to the synthetic phonics group and has a reading age higher than their chronological age at the second test.
- (c) The probability that a randomly selected child whose reading age is higher than their chronological age at the second test was initially allocated to the synthetic phonics group.
- (d) The probability that a randomly selected child has a reading age higher than their chronological age at the second test and was initially allocated to the synthetic phonics group.



'I know I'm having trouble reading at a 3rd grade level...that's why, when I grow up, I'm going to be a 2nd grade teacher.'

In Unit 6 (Subsection 2.3), you learned the multiplication rule for probabilities of statistically independent events: if events A and B are statistically independent, then

$$P(A \text{ and } B) = P(A) \times P(B).$$

The 'and' linkage leads to the *multiplication* of probabilities. A more general version of this rule can now be given, which applies also to pairs of events that are *not* statistically independent. The generalisation involves conditional probabilities.

Consider again the joint probability of events R_2 and AP , considered in Example 7. This may be written

$$\begin{aligned} P(R_2 \text{ and } AP) &= \frac{\text{number of children with event } AP \text{ and event } R_2}{\text{total number of children}} \\ &= \frac{\text{number of children with } AP}{\text{total number of children}} \\ &\quad \times \frac{\text{number of children with event } R_2 \text{ and event } AP}{\text{number of children with } AP}, \end{aligned}$$

because the terms 'number of children with AP ' cancel out top and bottom. Notice that we have

$$P(AP) = \frac{\text{number of children with } AP}{\text{total number of children}}$$

and

$$P(R_2|AP) = \frac{\text{number of children with event } R_2 \text{ and event } AP}{\text{number of children with } AP}.$$

So, substituting these expressions into the first equation, we have

$$P(R_2 \text{ and } AP) = P(AP) \times P(R_2|AP).$$

This result can be reached in other ways. For instance, suppose a randomly chosen child is to experience both AP and R_2 . Then:

1. The child must be in the AP group, which happens with probability $P(AP)$.
2. This child from the AP group must experience R_2 , which happens with probability $P(R_2|AP)$.

The probability that both 1 *and* 2 happen is $P(AP) \times P(R_2|AP)$, in keeping with multiplication being appropriate for the ‘and’ linkage.

More generally, for any two events A and B ,

$$P(A \text{ and } B) = P(B) \times P(A|B).$$

Note also that exactly the same argument yields

$$P(A \text{ and } B) = P(A) \times P(B|A),$$

since

$$P(A \text{ and } B) = P(B \text{ and } A).$$

These facts about joint and conditional probabilities are summarised in the following box.

Joint and conditional probabilities

Let A and B be any two events. The joint probability of A and B , denoted $P(A \text{ and } B)$, is the probability that both A and B occur.

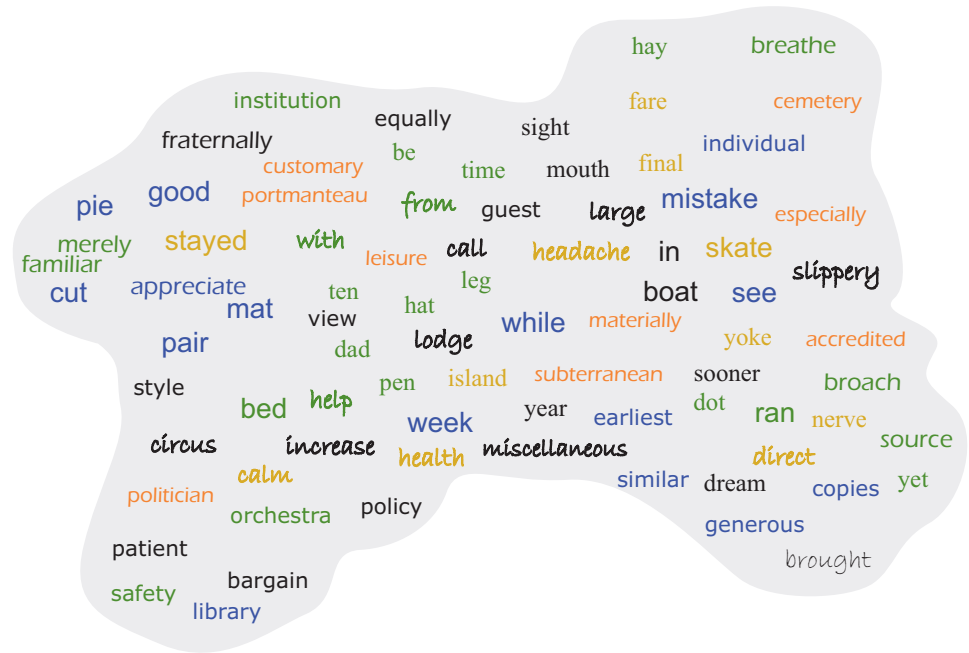
The conditional probability of A given B , denoted $P(A|B)$, is the probability that A occurs, given that B occurs.

Joint and conditional probabilities are linked by the following relationships:

$$P(A \text{ and } B) = P(A) \times P(B|A) = P(B) \times P(A|B).$$

Example 8 illustrates how these relationships work out on real data. For a change, we shall look at data other than reading data: this time, the data relate to spelling age, rather than reading age.

In the Clackmannanshire study, spelling age was measured using the Schonell Spelling Test, which, like the BAS reading test, is standardised to chronological age. As with the reading test, the spelling test was applied at baseline, prior to starting teaching. Then a first follow-up test was applied after the 16 school weeks during which the three teaching programmes were delivered. Following this test, all children were switched to the synthetic phonics programme. Finally, a second follow-up spelling test was applied at the end of the second school year.



Some of the words used in the Schonell Spelling Test

Example 8 *Spelling at baseline*

The data obtained in the spelling tests at baseline are in Table 13. The categories across the columns are similar to those for the reading test: ‘Higher’ in this case means ‘spelling age at baseline test higher than chronological age’.

Table 13 Results from the spelling test at baseline

	Spelling age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	57	51	108
Analytic phonics + PA	43	35	78
Synthetic phonics	70	47	117
Total	170	133	303

Consider the data from Table 13. Let S_0 denote the event ‘spelling age higher than chronological age at the baseline test’.

Then

$$P(AP) = \frac{108}{303} \simeq 0.356.$$

Also,

$$P(S_0|AP) = \frac{51}{108} \simeq 0.472.$$

Hence

$$\begin{aligned} P(AP) \times P(S_0|AP) &\simeq 0.356 \times 0.472 \\ &\simeq 0.168. \end{aligned}$$

Note that this equals $P(S_0 \text{ and } AP)$ [and $P(AP \text{ and } S_0)$]:

$$P(S_0 \text{ and } AP) = \frac{51}{303} \simeq 0.168.$$

Activity 9 will give you some further practice at manipulating joint and conditional probabilities.

Activity 9 *Calculating another joint probability*

In one school, 30% of sixth form students study physics. Of the students who study physics, 80% also study chemistry. If a student is picked at random, what is the probability that this student studies both physics and chemistry?

The relationship $P(A \text{ and } B) = P(B) \times P(A|B)$ can be rearranged so as to obtain

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}.$$

This provides a convenient way of obtaining the conditional probability $P(A|B)$ when the joint probability $P(A \text{ and } B)$ and the probability $P(B)$ are known. Activity 10 will give you some more practice at manipulating joint and conditional probabilities in this way.

Activity 10 *Comparing spelling at baseline*

Use the data in Table 13 as follows.

- Calculate $P(S_0|AP+)$ and $P(S_0|SP)$ directly from the table.
- Now derive these probabilities using the general relationship $P(A|B) = P(A \text{ and } B)/P(B)$, and check your results are the same as in part (a). (Make sure you keep full calculator accuracy for intermediate results.)
- Using these results and Example 8, comment briefly on any apparent differences in spelling ability at baseline.

You have now covered the material related to Screencast 1 for Unit 8 (see the M140 website).



3.2 Adding probabilities

In Unit 6 (Subsection 2.2), you learned how to add probabilities of mutually exclusive events: if events A and B are mutually exclusive, then

$$P(A \text{ or } B) = P(A) + P(B).$$

The ‘or’ linkage leads to the *addition* of probabilities. In this subsection, this expression will be extended to encompass events that are not mutually exclusive. (We shall find that subtraction can also be involved.)



In the classroom

Example 9 Spelling and reading at first follow-up

At the first follow-up in the Clackmannanshire study, after the children had received 16 school weeks of teaching according to the three programmes, they were tested again using the Schonell Spelling Test.

Table 14 shows the spelling test and reading test results together, using a new variable ‘Standardised spelling and reading ages at first follow-up test’. This variable has four categories – a similar variable was used in Table 6 (Subsection 2.2). The categories are Low/Low (spelling age and reading age both not higher than chronological age), Low/High (spelling age not higher and reading age higher than chronological age), High/Low (spelling age higher and reading age not higher than chronological age) and High/High (both spelling age and reading age higher than chronological age).

Table 14 Spelling and reading ability at first follow-up test

	Spelling/reading ages at first follow-up test				Total
	Low/Low	Low/High	High/Low	High/High	
Analytic phonics	65	14	3	22	104
Analytic phonics + PA	42	8	9	16	75
Synthetic phonics	29	7	6	71	113
Total	136	29	18	109	292

Let S_1 denote the event ‘spelling age higher than chronological age at first follow-up test’. Consider the totals across the teaching groups at the bottom of the table. A child did well in the spelling test if either the child is in the High/Low category or in the High/High category. So as $18 + 109 = 127$, that means 127 children out of the 292 did well in the spelling test (that is, their spelling age was higher than their chronological age), and so

$$P(S_1) = \frac{127}{292} \simeq 0.435.$$

A similar approach can be used for the different teaching groups. For example, consider the probability that a child is either in the analytic phonics (AP) group or in the analytic phonics + PA ($AP+$) group. As $104 + 75 = 179$, there are 179 children out of 292 in either the AP or $AP+$ group, so

$$P(AP \text{ or } AP+) = \frac{179}{292} \simeq 0.613.$$

Also,

$$P(AP) = \frac{104}{292} \simeq 0.356,$$

and

$$P(AP+) = \frac{75}{292} \simeq 0.257.$$

Thus $P(AP) + P(AP+) \simeq 0.613$. This equals $P(AP \text{ or } AP+)$, as expected, since the categories AP and $AP+$ are mutually exclusive.

In Example 9, we considered probabilities of mutually exclusive events. But now consider the probability of being in the High/High (HH) category or in the synthetic phonics (SP) group, that is, the probability $P(HH \text{ or } SP)$.

From Table 14, there are 109 children in the High/High group, and 113 children in the SP group. But the number in the ' HH or SP ' group does not equal $109 + 113$, because there are 71 children who are in both the HH group and the SP group: these children would be double-counted. The number of children who are in either the HH group or the SP group (or both) is

$$22 + 16 + 71 + 6 + 7 + 29 = 109 + 113 - 71 = 151,$$

out of the total 292 children in the table.

Hence

$$\begin{aligned} P(HH \text{ or } SP) &= P(HH) + P(SP) - P(HH \text{ and } SP) \\ &= \frac{109}{292} + \frac{113}{292} - \frac{71}{292} \\ &\simeq 0.517. \end{aligned}$$

This is an instance of the general addition rule, highlighted below.

The general addition rule for probabilities

Let A and B denote two events, mutually exclusive or not. Then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Activity 11 gives you some practice at using the general rule for adding probabilities.



Activity 11 Adding probabilities

- (a) In a school, 30% of sixth form students study physics, and 32% study chemistry. The proportion who study both physics and chemistry is 24%. What proportion of sixth form students in the school study physics or chemistry (or both)?
- (b) A total of 291 children in the study took both the baseline and the first follow-up reading tests. Of these, 110 did well at the baseline test (that is, their reading age was higher than their chronological age), and 138 did well at the first follow-up test. Sixty children did well in both tests. Calculate the proportion of children who did well on either the baseline test or the first follow-up test.
- (c) Among the children who did both the first and second follow-up reading tests, the proportions doing well (that is, with reading age higher than chronological age) were: 0.483 on the first test, 0.840 on the second test, and 0.848 on at least one of the tests. Calculate the proportion who did well on both tests. How many children among the 263 who did both tests did well on both?

3.3 Statistical independence and contingency tables

Recall from Unit 6 (Subsection 2.3) that two events are said to be *statistically independent* if the occurrence of one is unrelated to the chance of occurrence of the other.

In this subsection, we shall explore the notion of statistical independence as it applies to contingency tables. As before, we will use the Clackmannanshire study as an illustration.

Example 10 Spelling at first and second follow-ups

In Table 14 (Subsection 3.2) the joint results for spelling and reading at the first follow-up test were presented. We will now just concentrate on the spelling results. For ease of presentation, these are summarised in Table 15.

Table 15 Results from the spelling test at first follow-up

	Spelling age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	79	25	104
Analytic phonics + PA	50	25	75
Synthetic phonics	36	77	113
Total	165	127	292

The conditional probabilities of spelling age being higher than chronological age at the first follow-up are 0.240 (that is, 25/104) for the analytic phonics group, 0.333 (25/75) for the analytic phonics + PA group,

and 0.681 (77/113) for the synthetic phonics group. These values are different. Recall that in Activity 10 (Subsection 3.1) you found these probabilities were broadly similar at baseline. So as with the reading test, there had been a great improvement in the synthetic phonics group, but not in the other two groups. Thus, on the face of it, it would appear that the teaching method may have influenced the outcome of the first spelling test. In other words, spelling ability and teaching method may not be independent. (We have to hedge such statements with ‘appear that’ and ‘may not be’, as we need to allow for the possibility that the observed pattern is entirely due to chance. After Section 4, you will be able to make much stronger statements!)

Note also that the conditional probabilities for each teaching group differ from the overall probability that a child who took the test at first follow-up did well at spelling, which is

$$P(S_1) = \frac{127}{292} \simeq 0.435.$$

After the first follow-up test, all children in the study were taught using the synthetic phonics method, and tested again at the end of the following school year. In Example 6 (at the beginning of this section), it was suggested that the children in the two analytic phonics groups had caught up with those in the synthetic phonics group, in terms of their reading ability. So, what about their spelling ability? This is shown in Table 16.

Table 16 Results from the spelling test at second follow-up

	Spelling age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	10	85	95
Analytic phonics + PA	7	59	66
Synthetic phonics	10	94	104
Total	27	238	265

In Activity 12 you will interpret these results. Let S_2 denote the event ‘spelling age is higher than chronological age at the second-follow-up test’.

Activity 12 *Interpreting the second follow-up spelling test results*

- Using Table 16, obtain the conditional probabilities $P(S_2|AP)$, $P(S_2|AP+)$ and $P(S_2|SP)$.
- Obtain the probability $P(S_2)$, using the marginal totals. (Recall that marginal totals were defined in Example 3, Subsection 2.1.)
- Compare these probabilities. In your view, has the teaching group (to which the children were originally allocated) influenced the outcome of the second spelling test? How might you explain this, in the light of the results of the spelling test at first follow-up?



In Example 10 and Activity 12, we compared conditional and marginal probabilities and arrived at judgements about possible dependence of test results on study group on that basis. This makes good sense more generally. If two events A and B are statistically independent, then conditioning on event B will not affect the chance that A occurs. This also works the other way around: conditioning on A will not affect the chance that B occurs.

So far, we have discussed statistical independence in terms of pairs of *events*. However, in Activity 12, the interpretation of the results was really in terms of the relationship between the variables ‘teaching method’ and ‘spelling ability at the second test’; the events involved were just the values taken by those two variables. Thus, it makes sense to extend the notion of independence from pairs of events to pairs of *variables*: two variables are said to be statistically independent if there is no relationship between them, so that the value taken by either variable does not affect the value taken by the other variable. Thus, to say that the variables ‘teaching method’ and ‘spelling ability at the second test’ are independent means that the teaching method does not affect ability in the second follow-up spelling test, and vice versa.

Independence and conditional probabilities

Two events A and B are statistically independent if the occurrence of one has no influence on the chance of occurrence of the other.

Moreover, they are independent exactly when

$$P(A|B) = P(A)$$

or (equivalently)

$$P(B|A) = P(B).$$

Two *variables* are said to be independent if the value taken by either variable does not influence the value taken by the other variable.

Activity 13 will give you some practice at recognising whether variables are independent.

Activity 13 *Assessing independence*

In a school, 48% of sixth formers study both English and maths, 12% study maths but not English, 32% study English but not maths, and the remainder, 8%, study neither English nor maths. These percentages are shown as probabilities in Table 17.

Table 17 Proportions of students studying English and maths

	Studying maths	Not studying maths
Studying English	0.48	0.32
Not studying English	0.12	0.08

Let E denote the event ‘studies English’, and M the event ‘studies maths’.

- Obtain the probabilities $P(E \text{ and } M)$, $P(E)$ and $P(M)$.
- Hence obtain the conditional probabilities $P(E|M)$ and $P(M|E)$.
- What do you conclude about the independence or otherwise of studying English and studying maths?

You have now covered the material related to Screencast 2 for Unit 8 (see the M140 website).

Exercises on Section 3**Exercise 4** *Calculating probabilities from a contingency table*

Table 18 (which is the same as Table 9 in Exercises on Section 2) gives the numbers of boys and girls by teaching group in the Clackmannanshire study.

Table 18 Gender distribution by teaching group

	Gender		Total
	Boys	Girls	
Analytic phonics	58	51	109
Analytic phonics + PA	39	39	78
Synthetic phonics	61	56	117
Total	158	146	304

Let B denote the event ‘boy’ and G the event ‘girl’. Calculate the following probabilities for the children in the study.

- $P(G|AP)$
- $P(AP+)$
- $P(SP|B)$
- $P(B \text{ and } AP+)$

- (e) The probability that a child allocated to the synthetic phonics group is a girl.
- (f) The probability that a boy is allocated to the analytic phonics group.



Exercise 5 *Calculating more probabilities*

Of the 264 children who did both the spelling and the reading test at the second follow-up, 237 had a spelling age higher than chronological age (event S_2), whereas 220 had a reading age higher than chronological age (event R_2). Furthermore, 214 children had both spelling age and reading age higher than chronological age.

- (a) Calculate $P(R_2|S_2)$.
- (b) Calculate the probability that a child who did well at reading also did well at spelling.
- (c) Calculate $P(R_2 \text{ or } S_2)$.



Exercise 6 *Are reading and spelling independent?*

Table 19 shows the 2×2 contingency table of spelling and reading ability at the second follow-up test. To simplify the row and column headings, we have used 'Low' to indicate that the measured age (for reading or spelling) is not higher than chronological age, and 'High' to indicate that it is higher.

Table 19 Spelling and reading ability at second follow-up test

		Reading age		
		Low	High	Total
Spelling age	Low	21	6	27
	High	23	214	237
Total		44	220	264

- (a) Explain why Table 19 is a contingency table.
- (b) Obtain the probabilities $P(\text{Reading High}|\text{Spelling Low})$ and $P(\text{Reading High})$.
- (c) Compare these probabilities informally. Are the reading and spelling variables likely to be independent? What further information might you require?

4 The χ^2 test for contingency tables

At several points in Sections 2 and 3, you have been invited to comment about an association, or lack of association, between teaching method and reading (or spelling) ability. Thus, for example, you found in Activity 3 (Subsection 2.1) that the proportions of children doing well at reading in the baseline test were similar in the three reading groups; in Activity 4 (Subsection 2.1), in contrast, you found differences between groups in reading ability at the first follow-up test. But at what point does ‘similar’ become ‘different’? Any conclusions you might have been tempted to draw had to be qualified, owing to the possibility of chance fluctuations.

In this section, these issues are considered in detail. You will learn how to carry out the **χ^2 test for contingency tables**. (χ^2 is sometimes written as ‘chi-squared’ or ‘chi-square’ – particularly in situations where it is difficult to write the symbol ‘ χ^2 ’.) This test, first developed by Karl Pearson in 1900, provides a way of quantifying the likely effect of chance. We start by considering the hypotheses for the test.

Karl Pearson (1857–1936) is credited with founding the discipline of mathematical statistics. He also coined the term ‘contingency table’.

4.1 The null and alternative hypotheses

We shall consider how to formulate appropriate null and alternative hypotheses using some of the data encountered earlier. But first, it’s important to be clear that hypothesis tests cannot directly tell you *why* variables might be related, but only *whether* they are related. So, for example, hypothesis testing cannot tell you that the presence of one factor causes another to occur.



Karl Pearson (1857–1936)

Example 11 Taking care with hypotheses

Table 3 (Subsection 2.1) showed how the sample of children in the Clackmannanshire study performed at the first follow-up reading test. The question of interest is whether teaching method, represented by the categories determining the rows of the table, has any impact on reading ability, represented by the categories in the columns. So, it would seem that one way of expressing the null and alternative hypotheses could be:

H_0 : Teaching method has no effect on the reading ability at the first follow-up test of children starting at primary school.

H_1 : Teaching method does affect the reading ability at the first follow-up test of children starting at primary school.

As usual, the hypotheses relate to the whole population of children learning how to read, not to the ones that happened to be selected for the

The contrast between causality and association is discussed in Subsection 5.1.

sample. ‘Teaching method’, on the other hand, refers to the three methods investigated in the study.

This form of the alternative hypothesis implies that teaching method *causes* the child to score differently in the reading test. However, the hypothesis test that we develop does *not* actually investigate this. Rather, it examines whether teaching method and reading ability are associated. A more accurate way of expressing the hypotheses, which implies nothing about causality, will be discussed below.

In Activity 4 (Subsection 2.1), you calculated the percentages of children in each teaching group with a reading age higher than chronological age. The corresponding proportions are shown in Table 20.

Table 20 Results from the first follow-up test

	Reading age as compared to chronological age	
	Proportion not higher	Proportion higher
Analytic phonics	0.654	0.346
Analytic phonics + PA	0.680	0.320
Synthetic phonics	0.310	0.690

The proportions add up to 1 across rows, and vary as we read down each column of Table 20.

What we want to know is whether these differences are due to sampling or whether there are ‘real’ differences which are associated with the teaching method. If the study had been repeated in a different year, with different children, or choosing different schools, we would not expect the results to be exactly the same as in Table 20. However, we should expect them to be fairly similar, as the values were obtained from quite large samples and from a number of schools. What we want to do is to make an inference from the sample back to the population.

Thus, we are interested in the reading ability of the population of all children starting at primary school, and whether they would have the same distribution of reading abilities, whatever teaching method they had been taught by. In other words, we want to know whether teaching method and reading ability are independent.

This leads to the following form for the null and alternative hypotheses.

- H_0 : Teaching method for children starting at primary school and reading ability at the first follow-up test are independent.
- H_1 : Teaching method for children starting at primary school and reading ability at the first follow-up test are not independent.

This formulation of the hypotheses avoids any suggestion of causality. Activity 14 will give you some practice at formulating null and alternative hypotheses.

Activity 14 *Formulating hypotheses*

As discussed in Example 2 (Subsection 2.1), the baseline reading test was designed to find out whether there are any differences in reading ability between teaching groups before the different teaching methods were applied, because differences other than those due to chance could bias subsequent comparisons. Table 21 shows the observed proportions.

Table 21 Results from the baseline test

	Reading age as compared to chronological age	
	Proportion not higher	Proportion higher
Analytic phonics	0.639	0.361
Analytic phonics + PA	0.551	0.449
Synthetic phonics	0.650	0.350

Write down the null and alternative hypotheses to test whether there is any relationship between teaching method and reading ability at baseline.

4.2 Calculating the Expected values

In Unit 5 (Subsection 3.2) the DFR equation was introduced, which connects ‘Data’, ‘Fit’ and ‘Residual’. In this section, we shall use slightly different terminology, which is standard for contingency tables. Thus, ‘Data’ will be called ‘Observed’, and ‘Fit’ will be called ‘Expected’. This is because the ‘Fit’ values are those that one would expect if the null hypothesis of independence were true. The test to be described in this subsection is based on the following form of the DFR equation,

$$\text{Residual} = \text{Observed} - \text{Expected}.$$

The idea behind the test is as follows. If the null hypothesis is true, then the ‘Expected’ values should be close to the ‘Observed’ values, because these ‘Expected’ values are obtained assuming the null hypothesis is true. Thus, the residuals should be small if the null hypothesis is true. Consequently, if some of these residuals are large, it suggests that the null hypothesis is false. The χ^2 test statistic will be based on the sizes of the residuals.

The first step in calculating the χ^2 test statistic is to obtain the ‘Expected’ values. One way to do this is described in Example 12.

Example 12 *Expected values for reading at first follow-up*

We will use the contingency table of reading ability at the first follow-up test, given initially in Table 3 (Subsection 2.1) and reproduced below as Table 22.

Table 22 Results from the first follow-up test

	Reading age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	68	36	104
Analytic phonics + PA	51	24	75
Synthetic phonics	35	78	113
Total	154	138	292

We shall calculate the Expected values based on the null hypothesis of independence. This is as follows (for completeness we also state the alternative hypothesis):

H_0 : Teaching method for children starting at primary school and reading ability at the first follow-up test are independent.

H_1 : Teaching method for children starting at primary school and reading ability at the first follow-up test are not independent.

Under the null hypothesis of independence, the population distributions of reading ability are the same in each teaching group. So we can combine reading groups and use the column totals in Table 22 to calculate the proportions of children whose reading age is higher, or not higher, than chronological age.

Thus, under the null hypothesis,

$$P(\text{Reading age not higher than chronological age}) = \frac{154}{292},$$

and

$$P(\text{Reading age higher than chronological age}) = \frac{138}{292}.$$

We have not expressed these numbers as decimals as we are about to use them in calculations, and it's best to keep them as fractions. (You can of course use decimals, in which case make sure that you keep full calculator accuracy.)

Since there are 104 children in the analytic phonics group, then under the null hypothesis we would expect a proportion $154/292$ of them to have reading age not higher than chronological age, and a proportion $138/292$ to have reading age higher than chronological age.

Thus, the Expected values for the analytic phonics group are:

$$\text{Not higher: } 104 \times \frac{154}{292} \simeq 54.8493, \quad \text{Higher: } 104 \times \frac{138}{292} \simeq 49.1507.$$

Note that Expected values are not, in general, whole numbers. This is because they are not counts, but expected frequencies.

In Activity 15, you are invited to complete the Expected table using the method described in Example 12.

Activity 15 *Completing the Expected table*

Use the marginal counts in Table 22 to obtain the Expected values (to four decimal places):

- (a) for the analytic phonics + PA group
- (b) for the synthetic phonics group.



The Expected values obtained in Example 12 and Activity 15 are displayed in Table 23. (For brevity, the top heading of the results table is not included here. We will continue to use this abbreviated style for Expected and Residual tables.)

Table 23 Expected table for reading ability at first follow-up

	Not higher	Higher	Total
Analytic phonics	54.8493	49.1507	104.0000
Analytic phonics + PA	39.5548	35.4452	75.0000
Synthetic phonics	59.5959	53.4041	113.0000
Total	154.0000	138.0000	292.0000

Note that the marginal totals for the Expected table are the same as for the Observed table. This is no coincidence, but follows from the way the Expected values were calculated. (Occasionally some small discrepancies might occur, owing to rounding.)

The calculation of Expected values is summarised in the following box.

Calculating the Expected values

For each cell in a contingency table, the Expected values under the null hypothesis of independence are obtained as follows from the marginal totals:

$$\text{Expected value} = \frac{\text{row total} \times \text{column total}}{\text{overall total}}.$$

Activity 16 will give you some practice at calculating Expected values. A convenient way to do the calculations is to copy the marginal totals of the Observed table into the Expected table. Then calculate the Expected values from these marginal totals, and finally check that the Expected values add up to the marginal totals (apart from rounding error).



Activity 16 *Calculating Expected values for reading at baseline*

Obtain the Expected values (to four decimal places) for the contingency table of baseline reading test results in Table 2 (Subsection 2.1), reproduced here as Table 24. (This is the table of Observed values.)

Table 24 Observed results from the baseline test

	Reading age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	69	39	108
Analytic phonics + PA	43	35	78
Synthetic phonics	76	41	117
Total	188	115	303



You have now covered the material related to Screencast 3 for Unit 8 (see the M140 website).



Having waited in all day, Percy's mood was worsened even further when the furniture company finally turned up – and delivered a new sofa instead of the expected table.

4.3 Calculating the χ^2 test statistic

As stated at the beginning of Subsection 4.2, the χ^2 test statistic is based on the residuals. Having obtained the Expected values under the null hypothesis of independence, the next stage is to obtain the Residual table. This is described in Example 13.

Example 13 *Residuals for reading at first follow-up*

In Subsection 4.2, the Observed values are in Table 22, and the Expected values are in Table 23. The Residual table is obtained using the DFR equation in the form

$$\text{Residual} = \text{Observed} - \text{Expected}.$$

For example, the entry in the Residual table for the cell corresponding to ‘Not higher’ in the analytic phonics group is

$$68 - 54.8493 = 13.1507.$$

The other entries are calculated in a similar fashion, resulting in Table 25.

Table 25 Residual table for reading ability at first follow-up

	Not higher	Higher	Total
Analytic phonics	13.1507	−13.1507	0.0000
Analytic phonics + PA	11.4452	−11.4452	0.0000
Synthetic phonics	−24.5959	24.5959	0.0000
Total	0.0000	0.0000	0.0000

Note that the rows and the columns add to zero (subject to rounding). This follows from the fact that the Observed and Expected tables have the same marginal totals.

The test statistic is based not on the residuals directly, but on a measure of their magnitude relative to the Expected values, called the χ^2 **contributions**. For each cell we calculate the quantity

$$\chi^2 \text{ contribution} = \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}.$$

There are sound mathematical reasons for using this particular formula. For example, the reason we square the residuals in this way is that we are interested in their departure from zero, in either the positive or negative direction, both providing evidence against the null hypothesis. And the reason we divide by the Expected value is to allow for differences in the magnitudes of these Expected values. Note that the χ^2 contributions must always be positive, or zero: they cannot be negative.

Finally, the χ^2 test statistic is obtained by summing the χ^2 contributions. These steps of the calculation are illustrated in Example 14.

Example 14 Calculating χ^2 test statistic for reading at first follow-up

The Expected and Residual values in Table 23 (Subsection 4.2) and Table 25, respectively, are combined to obtain the contributions to χ^2 . Thus, for the cell corresponding to ‘Not higher’ in the analytic phonics group,

$$\begin{aligned}\chi^2 \text{ contribution} &= \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \\ &= \frac{\text{Residual}^2}{\text{Expected}} \\ &= \frac{13.1507^2}{54.8493} \\ &\simeq 3.1530.\end{aligned}$$

Repeating the calculation for all six cells results in the χ^2 contributions table (or χ^2 table, for short) in Table 26.

Table 26 χ^2 contributions table for reading ability at first follow-up

	Not higher	Higher
Analytic phonics	3.1530	3.5186
Analytic phonics + PA	3.3117	3.6956
Synthetic phonics	10.1510	11.3279

Finally, the χ^2 test statistic is obtained by summing the six χ^2 contributions:

$$\begin{aligned}\chi^2 &= 3.1530 + 3.5186 + 3.3117 + 3.6956 + 10.1510 + 11.3279 \\ &= 35.1578.\end{aligned}$$

The final value of the test statistic, rounded to three decimal places, is 35.158.

These steps of the calculation are summarised in the following box.

Calculating the χ^2 test statistic

From the Observed (O) and Expected (E) tables, obtain the residuals $O - E$, and hence the χ^2 contributions:

$$\chi^2 \text{ contribution} = \frac{(O - E)^2}{E}.$$

The χ^2 test statistic is the sum of the χ^2 contributions over all the cells in the contingency table:

$$\chi^2 = \sum \frac{(O - E)^2}{E}.$$

Activity 17 will give you some practice at these calculations.

Activity 17 Calculating χ^2 test statistic for reading at baseline



In Activity 16 (Subsection 4.2) you calculated the Expected values for the contingency table of reading ability at the baseline test. These values, along with the Observed marginal totals, are reproduced in Table 27.

Table 27 Expected values for the baseline test data

	Not higher	Higher	Total
Analytic phonics	67.0099	40.9901	108
Analytic phonics + PA	48.3960	29.6040	78
Synthetic phonics	72.5941	44.4059	117
Total	188	115	303

Use these values to obtain the following.

- (a) Obtain the χ^2 contributions, using the Observed values in Table 24.
- (b) Hence obtain the χ^2 test statistic.

You have now covered the material related to Screencast 4 for Unit 8 (see the M140 website).



4.4 Interpreting the χ^2 test statistic

The χ^2 test statistic was constructed in such a way that large positive values provide evidence against the null hypothesis of independence between the row and column variables whose categories define the contingency table. But how can we make this more precise?

In fact, if the null hypothesis is true, then the approximate probability distribution of the χ^2 statistic is known: it is a member of the χ^2 **family of distributions**. Choosing the right reference distribution among this family will be considered after the next example.

Example 15 *Interpreting χ^2 test statistic for reading at first follow-up*

From Example 14 (Subsection 4.3), the χ^2 test statistic for the 3×2 contingency table of teaching method by reading ability at the first follow-up test is 35.158. If the null hypothesis were true, this test statistic would have the particular χ^2 distribution shown in Figure 3.

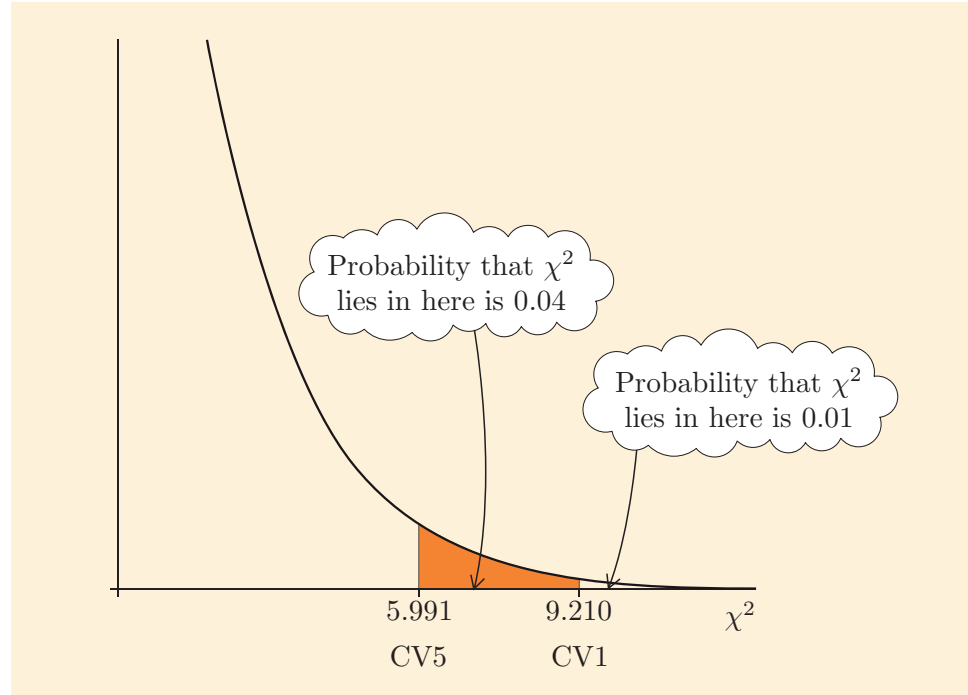


Figure 3 Probability distribution of the test statistic for reading ability at the first follow-up test, with critical values

Also shown in Figure 3 is the critical value CV5 corresponding to the 5% significance level, and the critical value CV1 corresponding to the 1% significance level. (Critical values were discussed in Units 6 and 7.) The 5% and 1% critical values are

$$CV5 = 5.991, \quad CV1 = 9.210.$$

The test procedure we use is very similar to that for the z -test described in Unit 7:

- Reject H_0 at the 1% significance level if $\chi^2 \geq 9.210$.
- Reject H_0 at the 5% significance level if $\chi^2 \geq 5.991$.
- Do not reject H_0 if $\chi^2 < 5.991$.

(If $5.991 \leq \chi^2 < 9.210$, then we reject H_0 at the 5% but not at the 1% significance level.)

Applying this rule, since in our case $\chi^2 = 35.158 \geq 9.210$, we reject the null hypothesis at the 1% significance level. To interpret this decision, we refer back to our hypotheses in Subsection 4.1. These were:

H_0 : Teaching method for children starting at primary school and reading ability at the first follow-up test are independent.

H_1 : Teaching method for children starting at primary school and reading ability at the first follow-up test are not independent.

We therefore conclude that the method used to teach children starting at primary school and their reading ability at the first follow-up test are not independent – that is, they are related.

Looking again at the table of χ^2 contributions (Table 26, Subsection 4.3), you will see that the biggest contributions to the χ^2 test statistic come from the synthetic phonics group. Now look at the signs of the residuals for the synthetic phonics group in the Residual table (Table 25, Subsection 4.3). The signs are negative for ‘Not higher’ and positive for ‘Higher’, indicating that, in the synthetic phonics group, there are fewer children whose reading age is not higher than chronological age, and more children whose reading age is higher, than would be the case under independence. Thus, the departure from independence appears to be mainly due to the higher ability of the synthetic phonics group.

Of course, you can also see this directly in this case: the data (first given in Table 3, Subsection 2.1) show that the proportion of children with reading age greater than chronological age is much higher in the synthetic phonics group than in the two other groups. However, with larger tables, it is sometimes more difficult to pin down where a departure from independence comes from, and in that case looking at the large χ^2 contributions and the signs of the corresponding residuals can help.

Example 15 shows that random fluctuations are unlikely to account for the better reading ability of the children in the synthetic phonics group at the first follow-up test. What about the baseline test? This is the topic of Activity 18.

Activity 18 *Interpreting χ^2 statistic for reading at baseline*

In Activity 17 (Subsection 4.3), you found that the χ^2 test statistic for the test of independence was 2.162. The critical values for the test are $CV_5 = 5.991$ and $CV_1 = 9.210$, as in Example 15.

Use this information to test the null hypothesis of independence between teaching group and reading ability in the baseline test. State your conclusions.

One final piece of the jigsaw remains to be put in place: choosing which member of the family of χ^2 distributions to use as reference for the test, and finding its critical values. The particular χ^2 distribution to use, and hence the critical values, depend on the **degrees of freedom** of the contingency table. These, in turn, depend on the size of the table.

For an $r \times c$ table, with r rows and c columns,

$$\text{degrees of freedom} = (r - 1) \times (c - 1).$$

For example, for a 3×2 table, $r = 3$ and $c = 2$, so

$$\text{degrees of freedom} = (3 - 1) \times (2 - 1) = 2.$$

Figure 4 shows the χ^2 probability distributions for several different values of the degrees of freedom.

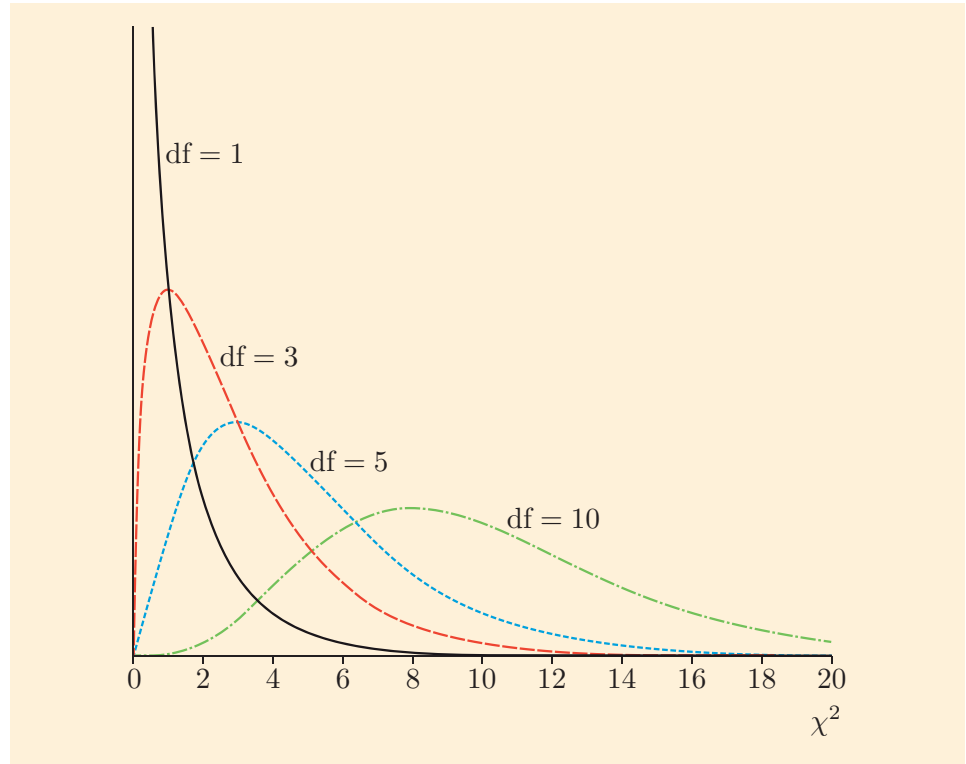


Figure 4 χ^2 probability distributions for degrees of freedom (df) 1, 3, 5, 10

Once you have worked out the degrees of freedom, you can look up the critical values CV5 and CV1 in statistical tables. A subset of the relevant table of critical values is given in Table 28.

Table 28 Table of critical values of χ^2

Degrees of freedom	Critical values of χ^2 at significance level	
	5%	1%
1	3.841	6.635
2	5.991	9.210
3	7.815	11.345
4	9.488	13.277
5	11.070	15.086
6	12.592	16.812
7	14.067	18.475
8	15.507	20.090
9	16.919	21.666
10	18.307	23.209
11	19.675	24.725
12	21.026	26.217

Activity 19 will give you some practice at finding critical values.

Activity 19 Finding critical values

Find the degrees of freedom and the critical values at significance levels 1% and 5% for contingency tables with the following sizes.

- (a) 2×2
- (b) 4×3
- (c) 5×4
- (d) 4×2

Tilting at windmills

When Karl Pearson first proposed the χ^2 test in 1900, he incorrectly specified that the degrees of freedom for an $r \times c$ contingency table should be $(r \times c) - 1$, rather than $(r - 1) \times (c - 1)$. For example, he would have given 8 rather than 4 as the degrees of freedom of a 3×3 contingency table. That all was not well was first noticed by statisticians Major Greenwood and Udny Yule in 1915, though they did not come up with a solution. This was provided in 1922 by the eminent statistician R. A. Fisher. Pearson did not accept he was wrong, and published a robust rebuttal in which he likened Fisher, whom he could not bring himself to name, to Don Quixote tilting at windmills. But Fisher was right: the correct number of degrees of freedom is $(r - 1) \times (c - 1)$.



Ronald Aylmer Fisher
(1890–1962)

You have now covered the material related to Screencast 5 for Unit 8 (see the M140 website).



4.5 Putting it all together

In this subsection we bring together the various steps involved in carrying out the χ^2 test. These are set out in the following box, and also in the flow diagram in Figure 5.

Procedure: the χ^2 test for contingency tables

1. Set up the null and alternative hypotheses in terms of the independence or otherwise of the row and column variables.
2. From the Observed table, calculate the Expected, Residual and χ^2 contribution tables.
3. Calculate the χ^2 test statistic by summing all the χ^2 contributions.
4. Find the degrees of freedom: $(r - 1) \times (c - 1)$. Look up the critical values at the 5% and 1% significance levels (CV5 and CV1).
5. Compare χ^2 with CV5 and CV1.
 - If $\chi^2 \geq \text{CV1}$, then H_0 is rejected at the 1% significance level.
 - If $\text{CV5} \leq \chi^2 < \text{CV1}$, then H_0 is rejected at the 5% significance level but not at the 1% significance level.
 - If $\chi^2 < \text{CV5}$, then H_0 is not rejected at the 5% significance level.
6. State your conclusion in non-mathematical terms.

As we saw after Example 15 (Subsection 4.4), if the null hypothesis is rejected, it can be helpful to look at the tables of χ^2 contributions and residuals to see in what way the two variables are not independent. First, look at the χ^2 contributions and pick out the large terms. These indicate which cells show a marked departure from independence. Then look at the residuals for these cells: their sign indicates the direction of the dependence.

In the remainder of this subsection, you will apply the χ^2 test to other results from the Clackmannanshire study. So far, you have seen that the baseline reading test results did not indicate any dependence between the teaching group and reading ability prior to teaching commencing. However, reading ability at the first follow-up test did show such a dependence (at the 1% significance level), and indicated that children in the synthetic phonics group were performing better.

In Activity 20 you will investigate whether this relationship persisted at the second follow-up test, after all children had been taught by the synthetic phonics method.

In this and subsequent activities and exercises, you should keep four decimal places in intermediate calculations, and round the χ^2 statistic to three decimal places.

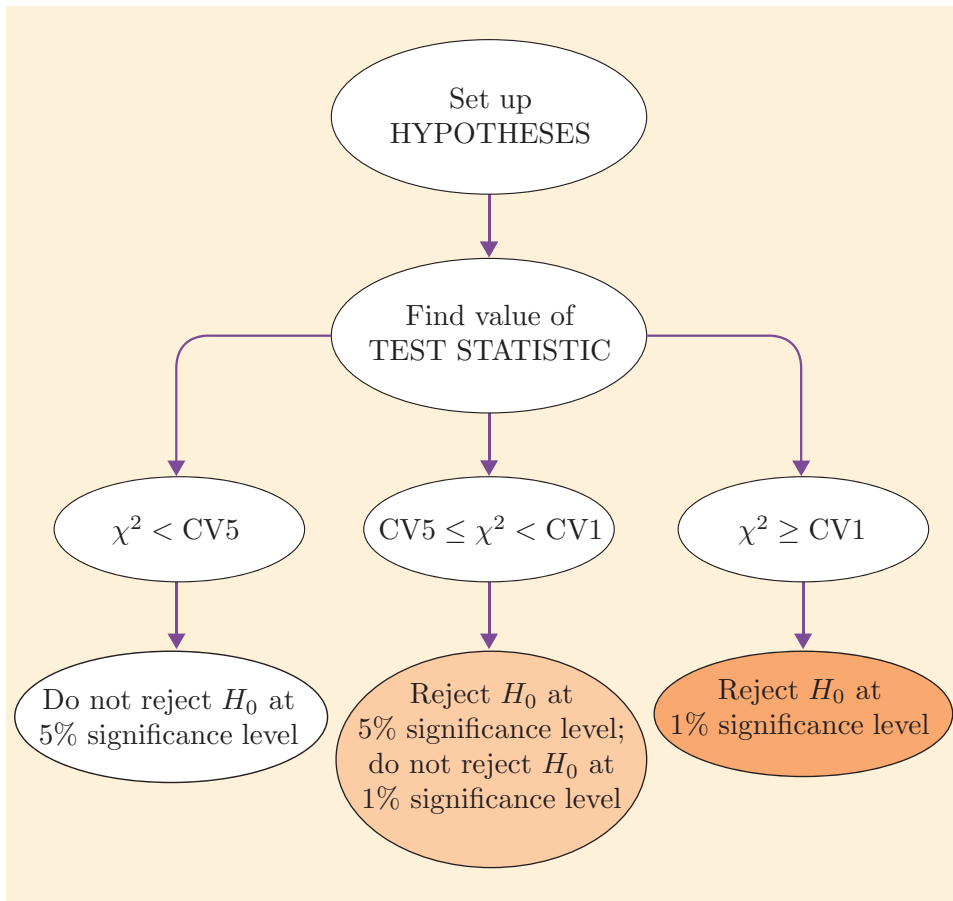


Figure 5 Flow diagram for the χ^2 test

Activity 20 *Testing for independence: reading at second follow-up*

The results of the second follow-up test were presented in Table 12 (Section 3), reproduced here as Table 29.



Table 29 Results from the second follow-up test

	Reading age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	19	78	97
Analytic phonics + PA	11	55	66
Synthetic phonics	15	90	105
Total	45	223	268

- Write down the null and alternative hypotheses (the two variables are: teaching method originally allocated, and reading ability).
- Obtain the Expected table.
- Hence obtain the Residual and χ^2 contributions tables.

- (d) Calculate the χ^2 test statistic and note the appropriate critical values, CV5 and CV1 (use Table 28 in Subsection 4.4).
- (e) State your conclusions.

Thus, by the second follow-up test, the difference between the teaching groups has disappeared. Looking at Table 29, it is clear that by the second follow-up test, the three groups of children are performing equally well, and a large proportion of children in each group now have a reading age higher than their chronological age. One interpretation is that switching the children in the two analytic phonics groups to synthetic phonics has allowed them to catch up.

A restriction of the χ^2 test, which we have not mentioned so far, is that it should not be used when some of the Expected values are small. This is because the calculation of the critical values is based on approximations which hold only when the Expected values are reasonably large. In M140, we shall use the following rule:

The Expected ≥ 5 rule

The χ^2 test should be used *only* when all Expected values are *greater than or equal to 5*.

The number 5 in this rule is a little arbitrary. There are other rules that are less restrictive, but this is the most commonly used one, so it makes sense to use it for M140. The rule applies only to the Expected values: the χ^2 test still applies if the Observed values are small.

Example 16 gives an example where the rule is violated, and describes a way round the problem by combining categories.

Example 16 *Revisiting spelling and reading at first follow-up*

In Example 9 (Subsection 3.2) data were presented on ability in both spelling and reading at the first test. The data from Table 14, introduced in that example, are reproduced here as Table 30.

Table 30 Spelling and reading ability at first follow-up test

	Spelling/reading ages at first follow-up test				Total
	Low/Low	Low/High	High/Low	High/High	
Analytic phonics	65	14	3	22	104
Analytic phonics + PA	42	8	9	16	75
Synthetic phonics	29	7	6	71	113
Total	136	29	18	109	292

The corresponding table of Expected values is shown in Table 31.

Table 31 Expected values for spelling and reading ability at first follow-up

	Low/Low	Low/High	High/Low	High/High
Analytic phonics	48.4384	10.3288	6.4110	38.8219
Analytic phonics + PA	34.9315	7.4486	4.6233	27.9966
Synthetic phonics	52.6301	11.2226	6.9658	42.1815

The Expected value corresponding to the High/Low category within the analytic phonics + PA group is 4.6233, and thus violates the rule. Hence, the results of the χ^2 test applied to this table may be unreliable.

One way to get round this problem is to combine categories. There is no general way to do this: some combinations are sensible, others less so. In the present case, we could combine the analytic phonics and analytic phonics + PA groups into a single ‘any analytic phonics’ group, so as to retain the contrast between methods based on analytic and synthetic phonics. (An alternative combination is explored in Exercise 9.)

Combining categories should only be done if it makes sense. If the contingency table is of size 2×2 , then combining categories is not possible. For example, if the two rows were combined you would no longer have a contingency table. (There is a test which you can apply in these circumstances, but it is not covered in M140.)

In Activity 21 you are asked to apply the χ^2 test to Table 30 with the two analytic phonics groups combined. This will produce a 2×4 table. The χ^2 test applies to tables of any dimensions $r \geq 2$, $c \geq 2$.

Activity 21 *Testing for independence: using combined groups*



In Example 16 it was suggested that the two groups ‘analytic phonics’ and ‘analytic phonics + PA’ should be combined into a new teaching group, ‘any analytic phonics’. These data are shown in Table 32.

Table 32 Spelling and reading ability at first follow-up test

	Spelling/reading ages at first follow-up test				Total
	Low/Low	Low/High	High/Low	High/High	
Any analytic phonics	107	22	12	38	179
Synthetic phonics	29	7	6	71	113
Total	136	29	18	109	292

- Write down the null and alternative hypotheses (the two variables are: teaching method, and ability in spelling/reading).
- Obtain the Expected table. Check that all Expected values are ≥ 5 .
- Obtain the Residual and χ^2 contributions tables.

- (d) Calculate the χ^2 test statistic and note the appropriate critical values, CV5 and CV1 (use Table 28 in Subsection 4.4).
- (e) State your conclusions.
- (f) If you reject the null hypothesis, describe the relationship you observe.

In this section we have developed the χ^2 hypothesis test for contingency tables, and applied it to data from the Clackmannanshire study. In the next section we shall look a little more closely at causality, and also consider some reservations about both the data and the hypothesis test.



You have now covered the material related to Screencast 6 for Unit 8 (see the M140 website).

Exercises on Section 4



Exercise 7 Testing for independence: spelling at first follow-up

The results on spelling ability at the first follow-up test were presented in Table 15 (Subsection 3.3), reproduced here as Table 33.

Table 33 Results from the spelling test at first follow-up

	Spelling age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	79	25	104
Analytic phonics + PA	50	25	75
Synthetic phonics	36	77	113
Total	165	127	292

- (a) Write down the null and alternative hypotheses.
- (b) Obtain the Expected table.
- (c) Hence obtain the Residual and χ^2 contribution tables.
- (d) Calculate the χ^2 test statistic and note the appropriate critical values, CV5 and CV1.
- (e) State your conclusions.
- (f) If you reject the null hypothesis, describe the relationship you observe.

Exercise 8 *Testing for independence: spelling at second follow-up*

The results on spelling ability at the second follow-up test were presented in Table 16 (Subsection 3.3), reproduced here as Table 34.

Table 34 Results from the spelling test at second follow-up

	Spelling age as compared to chronological age		
	Not higher	Higher	Total
Analytic phonics	10	85	95
Analytic phonics + PA	7	59	66
Synthetic phonics	10	94	104
Total	27	238	265

- Write down the null and alternative hypotheses. (The variables are: teaching method originally allocated, and spelling ability at the second follow-up test.)
- Obtain the Expected table.
- Hence obtain the Residual and χ^2 contributions tables.
- Calculate the χ^2 test statistic and note the appropriate critical values, CV5 and CV1.
- State your conclusions.

Exercise 9 *Spelling and reading: combining categories*

In Example 16 (Subsection 4.5), it was suggested that the two analytic phonics groups in Table 30 should be combined in order to make all Expected values greater than 5; you performed the corresponding χ^2 test in Activity 21. An alternative is to combine the Low/High and High/Low categories, thus resulting in the data shown in Table 35.

Table 35 Spelling and reading ability at first follow-up test

	Spelling/reading ages at first follow-up test			Total
	Low/Low	Low/High or High/Low	High/High	
Analytic phonics	65	17	22	104
Analytic phonics + PA	42	17	16	75
Synthetic phonics	29	13	71	113
Total	136	47	109	292

- Write down the null and alternative hypotheses. (The variables are: teaching method, and spelling/reading ability at the first follow-up test.)
- Obtain the Expected table. Check that the Expected values are all greater than or equal to 5.

- (c) Hence obtain the Residual and χ^2 contributions tables.
 - (d) Calculate the χ^2 test statistic and note the appropriate critical values, CV5 and CV1.
 - (e) State your conclusions.
 - (f) If you reject the null hypothesis, describe the relationship you observe.
-

5 More about the χ^2 test

In this section, some further aspects of the χ^2 test are considered. In fact, much of the material in this section applies more generally to hypothesis tests other than the χ^2 test. However, we shall use the χ^2 test as the focus of the discussion and, specifically, its application to the theme of this unit, namely the evaluation of different methods for teaching how to read.

The story so far may be summarised as follows. The Clackmannanshire study was undertaken to compare three methods for teaching how to read: analytic phonics, analytic phonics plus phonological awareness (PA), and synthetic phonics. To this end, 13 classes of children entering primary schools in Clackmannanshire were assigned to one of the three study methods. Before the teaching began, the children were tested on a range of abilities, including reading and spelling. At this stage, there were no significant differences between the three groups (as demonstrated by a χ^2 test at the 5% significance level). The children were retested after 16 teaching weeks. At this stage, there were significant differences in reading and spelling ability between the three groups (as shown by a χ^2 test at the 1% level), the children allocated to the synthetic phonics programme having performed much better than those on the two analytic phonics programmes. In consequence, all children were thereafter taught using synthetic phonics. A second follow-up test was done at the end of the second school year. A further χ^2 test on these data (at the 5% level) revealed no differences between the three groups: all three performed equally well. Our interpretation (see comment after Activity 20 in Subsection 4.5) was that the children originally allocated to the two analytic phonics groups had ‘caught up’.

This, however, is but *one* interpretation of the findings. The χ^2 test does not tell us how to interpret these findings. It can only tell us that, at the second follow-up, there is no significant difference between the three groups. It does not tell us *why* there is no difference, just as it did not tell us *why* a difference had appeared at the first follow-up test.

5.1 Causality and association

Of course, what we really want to know is whether synthetic phonics is a better method of teaching, that is, whether it *helps* children to learn how to read better than the other two methods. But no hypothesis test can, on its own, demonstrate causality or the lack of it. It can only provide evidence of a relationship or association between variables, and cannot show that a change in one variable *causes* a change in the other.

For example, another interpretation of the results at the second follow-up test could be that the children in the analytic phonics group would have ‘caught up’ anyway, whatever the teaching method. Under this interpretation, the early advantage of the synthetic phonics group at the first follow-up test does not persist: in the long term there is no difference between the methods.

In order to investigate these and other issues, a further comparative study was undertaken to compare the children (now aged 10) taught primarily by synthetic phonics in the original Clackmannanshire study, with an otherwise similar group of children of the same age who had been taught primarily using analytic phonics. The data in this subsection come from this further study, which we shall call the Long Term study. (Johnston, R., McGeown, S. and Watson, J. (2012) ‘Long-term effects of synthetic versus analytic phonics teaching on the reading and spelling ability of 10 year old boys and girls’, *Reading and Writing: An Interdisciplinary Journal*, vol. 25, pp. 1365–84.)

Owing to the widespread influence of the Clackmannanshire study, which led to synthetic phonics being widely adopted in Scottish schools, it was not possible to find a suitable comparison group in Scotland who had been taught by analytic phonics. Instead, a sample of classes in an English city was selected, who had been taught using a mixed approach including analytic phonics. The two samples are thus the 190 children aged 10 from the original Clackmannanshire study, who were taught primarily using synthetic phonics (which we shall call the ‘synthetic phonics’ group), and a comparable group of 203 children from schools in England, who were taught primarily using analytic phonics (the ‘analytic phonics’ group).

The reading test used was the Wide Range Achievement Test (WRAT test), while spelling was tested using the Schonell Spelling Test, as before. Both tests are standardised for chronological age. (The WRAT test has a higher upper age limit than the BAS test previously used.) Other tests were performed, notably reading comprehension, with similar results as for reading and spelling.



Activity 22 *Comparing reading abilities of 10-year-old children*

Table 36 shows the results of the WRAT reading test in the two groups of the Long Term study. As before, the reading ability variable has two categories: ‘Not Higher’ corresponds to reading age less than or equal to chronological age, while ‘Higher’ corresponds to reading age higher than chronological age. The age standardisation corrects for small age differences between children.

Table 36 Reading abilities at age 10

	Standardised reading age		
	Not higher	Higher	Total
Analytic phonics	115	88	203
Synthetic phonics	50	140	190
Total	165	228	393

- (a) Carry out the χ^2 test of independence between teaching group and reading ability.
- (b) What result would have been expected if there is no difference in the long term between the two teaching methods? What does this test suggest about the validity of this hypothesis?

In Activity 22, you found that the children in the synthetic phonics group did better at reading than those in the analytic phonics group, which suggests that the improvements in reading ability seen in the Clackmannanshire study are lasting. However, before reaching this conclusion, further issues have to be considered.

One issue is whether the two groups in the Long Term study are comparable with respect to other factors (besides teaching method) that might influence reading ability. For example, there is evidence that the socio-economic background of the child may have an impact on reading ability, with children from deprived backgrounds doing less well than children from advantaged backgrounds. Thus the observed association might have been induced by a difference in socio-economic backgrounds between the two groups, rather than by different teaching methods. This phenomenon, called **confounding**, is illustrated in Figure 6.

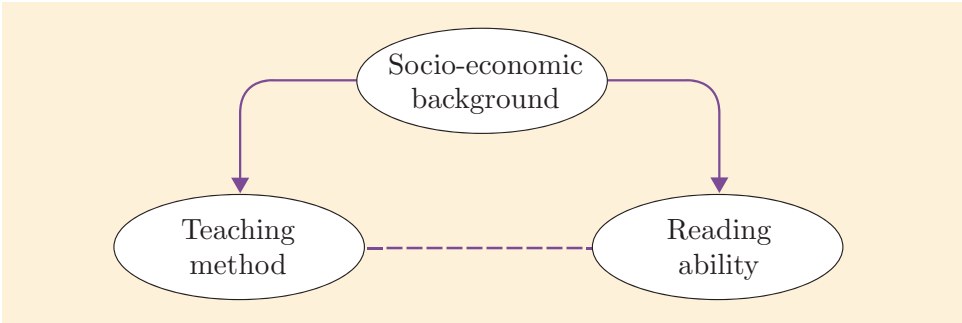


Figure 6 Confounding by socio-economic background

The arrows from ‘Socio-economic background’ to ‘Teaching method’ and from ‘Socio-economic background’ to ‘Reading ability’ indicate the presence of an association; the dashed line between ‘Teaching method’ and ‘Reading ability’ refers to an apparent association, which is actually induced by the other two: in this way, the true relationship (or lack of one) between ‘Teaching method’ and ‘Reading ability’ is *confounded* by socio-economic background.

Can such an effect explain the results observed in Table 36? This is the topic of Example 17.

Example 17 *The impact of socio-economic background on reading*

We want to know whether differences in socio-economic background can account for the pattern observed in Table 36. The children in the study were classified into two broad socio-economic groups, labelled ‘Advantaged’ and ‘Disadvantaged’. One way to remove any confounding effect of socio-economic group is to do two analyses: one with children from advantaged backgrounds and another with children from disadvantaged backgrounds. Since the children in each group have similar socio-economic backgrounds, each of the two analyses should be free of any confounding effects due to this variable.

Table 37 shows the data for children from advantaged backgrounds only.

Table 37 Reading abilities of 10-year-old children from advantaged backgrounds

	Standardised reading age		
	Not higher	Higher	Total
Analytic phonics	56	52	108
Synthetic phonics	21	82	103
Total	77	134	211

For these data, the test statistic is $\chi^2 = 22.520$. The Observed table is 2×2 , so has one degree of freedom, and so $CV_5 = 3.841$ and $CV_1 = 6.635$. (You could check these numbers for yourself.) Since $22.520 \geq 6.635$, we reject the null hypothesis of independence at the 1% significance level.

Examination of Table 37 shows that, for children from advantaged backgrounds, better results are obtained in the synthetic phonics group. What of children from disadvantaged backgrounds? Table 38 shows the data for children in this group.

Table 38 Reading abilities of 10-year-old children from disadvantaged backgrounds

	Standardised reading age		
	Not higher	Higher	Total
Analytic phonics	59	36	95
Synthetic phonics	29	58	87
Total	88	94	182

In this case, $\chi^2 = 15.054$. The critical values are the same as before. Since $15.054 \geq 6.635$, the null hypothesis is rejected at the 1% significance level. And again, examination of Table 38 shows that children in the synthetic phonics group do better than those in the analytic phonics group.

By doing separate analyses for children from advantaged backgrounds and children from disadvantaged backgrounds, we have removed the effect of socio-economic background. In both groups, children taught by synthetic phonics appear to do better than children taught by analytic phonics. Thus, the observed relationship between teaching method and reading ability is not the result of confounding by socio-economic background.



Confusion worse confounded

Jones: 'Con-found it all! Somebody's taken my hat, and left this filthy, beastly, shabby old thing instead!'
Brown: 'A-I beg your pardon, but that happens to be my hat!'

Ruling out socio-economic background as a potential confounder is a big step towards being able to assert that synthetic phonics really does work better as a teaching method than analytic phonics. However, there are other confounders to think about – notably gender. Girls and boys may respond differently to different teaching methods, so it is important to study the possible confounding effect of gender in the Long Term study. This is the topic of Activity 23.

Activity 23 *Investigating the impact of gender on reading*



Table 39 shows the data in the Long Term Study for boys only.

Table 39 Reading abilities of 10-year-old boys

	Standardised reading age		
	Not higher	Higher	Total
Analytic phonics	62	47	109
Synthetic phonics	20	86	106
Total	82	133	215

Table 40 shows the corresponding data for girls.

Table 40 Reading abilities of 10-year-old girls

	Standardised reading age		
	Not higher	Higher	Total
Analytic phonics	53	41	94
Synthetic phonics	30	54	84
Total	83	95	178

- (a) Perform the χ^2 test for independence in boys only using the data in Table 39.
- (b) Perform the χ^2 test for independence in girls only using the data in Table 40.
- (c) Interpret your findings in terms of the presence or otherwise of confounding by gender in the Long Term study.

In Activity 23 you found that both boys' and girls' reading abilities are better in the synthetic phonics groups. In fact, you can say a little more than that. This is the topic of Example 18.

Example 18 *Comparing boys and girls*

Using AP and SP to denote membership of the analytic phonics and synthetic phonics groups respectively, we can calculate the conditional probabilities of achieving ‘Higher’ in the two groups, for boys and girls separately. Thus, for boys, using Table 39 we obtain:

$$P(\text{Higher}|AP) = \frac{47}{109} \simeq 0.431,$$

$$P(\text{Higher}|SP) = \frac{86}{106} \simeq 0.811.$$

Similarly for girls, from Table 40,

$$P(\text{Higher}|AP) = \frac{41}{94} \simeq 0.436,$$

$$P(\text{Higher}|SP) = \frac{54}{84} \simeq 0.643.$$

Thus, girls and boys in the analytic phonics groups perform much the same: the percentage whose reading age is higher than chronological age is about 43% for boys and 44% for girls.

Both boys and girls do better in the synthetic phonics groups: the percentages doing well are about 81% for boys and 64% for girls. But note that boys do particularly well. The improvement in ability is about 38% for boys $[(0.811 - 0.431) \times 100\%]$ compared to 21% for girls $[(0.643 - 0.436) \times 100\%]$.

Thus, it appears that, while both girls and boys do better when taught using synthetic rather than analytic phonics, boys respond particularly well to that method of teaching.

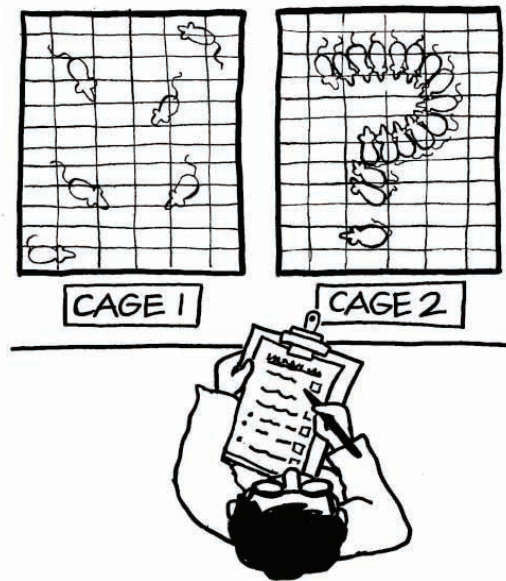
5.2 Reservations

We have spent a lot of time looking at data collected as part of the Clackmannanshire and Long Term studies. In this subsection, we critically consider the conclusions we drew from these studies.



*'I have a Ph.D. and 20 years experience.
But I'm a Statistician, so my conclusions
will always be more qualified than I am.'*

While both studies were carefully conducted, with due attention to comparability of groups, and outcomes assessed with well validated standardised tests, the children in each of the two studies were not actually selected at random from the populations of all children entering primary school, or of all 10-year-old primary school children. This, it must be said, is a feature of most studies, and does not in itself invalidate the results, but is something that must always be borne in mind. It is important to be on the lookout for any features of the data or of the study design that may suggest the individuals included or the measurements taken might be atypical in some way. In our case, there is nothing to suggest that the children and schools selected in these studies were atypical in any way.



Despite the clear difference between the treated and control groups, something made him question the data.

We now turn to reservations about our conclusions. There is one very important type of reservation that applies not only to the conclusions we have drawn in this unit but to all hypothesis tests. In a hypothesis test the procedure is to infer back from the sample to the appropriate population. The test used depends on the nature of the sample data and the question being posed, but in every case a similar procedure is adopted. We set up null and alternative hypotheses, find a test statistic from the sample data and compare it with critical values at the 5% and 1% significance levels to see if it lies in both, one or neither of the critical regions. If the test statistic lies in the critical region at the 1% significance level, then we reject the null hypothesis in favour of the alternative hypothesis at the 1% significance level and feel confident of our conclusion. If it lies in the critical region at the 5% significance level but not in that at the 1% significance level, we reject the null hypothesis at the 5% significance level and conclude that there is some evidence that the null hypothesis is false. If the test statistic lies in neither critical region then we conclude that there is insufficient evidence to reject the null hypothesis.

The use of a hypothesis test always carries with it the possibility of error, because we are using a sample to make inferences about a whole population. Even if the null hypothesis is true, random variation means that the test statistic obtained from a sample will sometimes unfortunately lie in the critical region. If we observed such a sample, we would reject the null hypothesis wrongly and so make an error. This type of error, where we reject the null hypothesis even though it is true, is known as a **type 1 error**.

Suppose that the null hypothesis (H_0) is true. If we use a significance level of 5%, there is a probability of 0.05 (5%) that the test statistic from a sample will lie in the critical region for the 5% significance level. That is, the probability of rejecting H_0 at the 5% significance level when H_0 is true is 0.05. Thus the probability of making a type 1 error is 0.05 if we use a 5% significance level. Similarly, if we reject H_0 at the 1% significance level, then when H_0 is true, the probability of making a type 1 error is 0.01. In general, the probability of making a type 1 error is given by the significance level at which H_0 is rejected.

It is also possible to make an error by *not* rejecting the null hypothesis when it is actually false. This is known as a **type 2 error**. It occurs when the particular sample we have picked has a test statistic that falls outside the critical region, even though the null hypothesis is false. Table 41 shows how the type 1 and type 2 errors relate to the null hypothesis.

Table 41 Type 1 and type 2 errors

	H_0 true	H_0 false
H_0 not rejected	Correct	Type 2 error
H_0 rejected	Type 1 error	Correct

You might be tempted to think that it is a good idea to make the probability of making a type 1 error as small as possible. However, for a given dataset, the two types of error are related, in that the smaller the probability of making a type 1 error, the larger the probability of making a type 2 error. So it is a matter of compromise, and the accepted procedure is to fix the probability of making a type 1 error.

The appropriate value to fix for this probability will vary with the situation. Most commonly, a probability of 0.05 of making a type 1 error is chosen. Sometimes, though, the action to be taken if the null hypothesis is rejected might be very expensive; in such a case, the probability of making a type 1 error might be fixed at 0.01 or even 0.001. As a result of such a choice, the probability of making a type 2 error would be larger.

It is not usually possible to quantify exactly the probability of making a type 2 error – the probability depends on the actual true form of the alternative hypothesis. Nevertheless, it is important to be aware of the possibility of making this type of error. The best way of reducing the probabilities of both types of error is to increase the sample size, by conducting a larger experiment.

Activity 24 Which type of error?

In each of the following scenarios, the ‘truth’ is first set out, then the result of a relevant hypothesis test is given. In each scenario, say whether a type 1 error, a type 2 error or no error has been made.

- (a) Reading ability after one year’s primary school teaching is related to teaching method; the null hypothesis of independence was not rejected.
- (b) Reading ability at age 10 is related to teaching method; the null hypothesis of independence was rejected.
- (c) Spelling ability after one year’s primary school teaching is unrelated to teaching method; the null hypothesis of independence was rejected.
- (d) Spelling ability at age 10 is unrelated to teaching method; the null hypothesis of independence was not rejected.

Exercises on Section 5

Throughout this unit, we have focused on data about learning how to read. The methods we have described are, of course, universally applicable. The following two exercises are drawn from different fields of application.



Exercise 10 Investigating gender differences in sandflies

The data in Table 42 show the numbers of male and female sandflies caught in two light-traps, one set 3 feet and the other 35 feet above the ground. Is there any evidence to suggest that the variables height and gender are related?

Table 42 Trapped sandflies

	Height above ground	
	3 feet	35 feet
Males	173	125
Females	150	73

(Data source: Hand, D.J. et al. (1994) *A Handbook of Small Data Sets*, London, Chapman & Hall)

- (a) Obtain the marginal totals, and set up the null and alternative hypotheses.
- (b) Carry out a χ^2 test and state your conclusion.
- (c) What type of error might you have committed?



A sandfly takes off: how high will it fly?

Exercise 11 *Investigating blood groups in Iraq*

Some data that were obtained from Iraq show the blood groups of samples of individuals from certain population groups.

**Table 43** Blood groups in Iraq

Population group	Blood group				Total
	O	A	B	AB	
Kurd	531	450	293	226	1500
Arab	174	150	133	36	493
Other	139	134	74	33	380
Total	844	734	500	295	2373

(Data source: Rohatgi, V.K. and Saleh, A.K. (2001) *An Introduction to Probability and Statistics*, 2nd edn, Wiley)

- Set up the null and alternative hypotheses appropriate to this contingency table.
- Carry out a χ^2 test to investigate whether there is any relationship between blood group and ethnic group.
- State whether you may have made a type 1 or a type 2 error in your conclusion. If you conclude that there is a relationship, indicate the main differences in the blood group distributions for the different population groups.

6 Computer work: the χ^2 test for contingency tables



In this section, you will learn how to use Minitab to perform the χ^2 test on contingency tables, and how to enter data from contingency tables for analysis.

You should now turn to the Computer Book and work through Chapter 8.

Summary

This unit has focused on the analysis of count data on two categorical variables, organised as a contingency table. Using data on the impact of different methods of teaching how to read, you learned to recognise contingency tables and calculate proportions from them.

The concepts of joint and conditional probabilities were described in terms of contingency tables. These types of probability are related by $P(A \text{ and } B) = P(A) \times P(B|A) = P(B) \times P(A|B)$, and they can be calculated from contingency tables. The addition rule for probabilities was extended to pairs of events that are not mutually exclusive. Also, the concept of statistical independence was related to conditional probabilities.

You learned how to formulate the null and alternative hypotheses for the χ^2 test for contingency tables, how to calculate the χ^2 test statistic, obtain the appropriate degrees of freedom, and obtain critical values for the test. You also learned how to interpret the results of the test in terms of the variables involved.

The distinction between association and causality was emphasised. The concept of confounding was described, along with how to investigate confounding by a third categorical variable. You were then introduced to type 1 and type 2 errors, and when they may occur.

Finally, you learned how to input contingency table data and how to carry out the χ^2 test in Minitab.

Learning outcomes

After working through this unit, you should be able to:

- recognise categorical data and categorical variables
- recognise a contingency table
- obtain the marginal totals of a contingency table
- calculate proportions from a contingency table
- state the dimensions of a contingency table
- recognise joint and conditional probabilities
- calculate joint and conditional probabilities from a contingency table
- state and use the relationship between joint and conditional probabilities
- state and use the general addition rule for probabilities
- express the independence of two events in terms of conditional probabilities
- carry out the χ^2 test for contingency tables, taking account of the size of Expected values
- interpret a χ^2 test in terms of the null and alternative hypotheses
- appreciate the distinction between association and causality
- investigate confounding by splitting a contingency table into several component tables and testing each table separately
- understand the concepts of type 1 and type 2 errors, and know when they may occur
- decide when to use the χ^2 test for contingency tables
- input contingency tables into Minitab
- use Minitab to carry out the χ^2 test.

Solutions to activities

Solution to Activity 1

The question starts with ‘what is the best way’. So one aspect to clarify is which methods are to be compared, and then some criterion for measuring ‘best’ will be required. The rest of the question is ‘teaching how to read’, so we will need to specify exactly what is meant by ‘how to read’.

Solution to Activity 2

A sensible precaution is to test the children’s reading ability before the different teaching methods are applied, in order to check that the children are starting from similar baselines. While differences in socio-economic or other factors between groups may still influence the effectiveness of the different teaching methods, demonstrating that the three groups are comparable from the start provides reassurance that the study is not fundamentally biased.

Solution to Activity 3

- (a) The proportion in the analytic phonics + PA group (denoted AP + PA) is

$$\frac{35}{78} \simeq 0.449,$$

or about 44.9%. In the synthetic phonics group (denoted SP), the proportion is

$$\frac{41}{117} \simeq 0.350,$$

or about 35.0%.

- (b) The proportions in the three groups are: AP: 36%; AP + PA: 45%; and SP: 35%. The analytic phonics group and the synthetic phonics group are very similar, but the analytic phonics + PA group starts with perhaps a slight advantage.

Solution to Activity 4

- (a) The proportion of children in the analytic phonics group (AP) with reading age higher than chronological age is

$$\frac{36}{104} \simeq 0.346,$$

or about 34.6%. The proportion in the analytic phonics + PA group (AP + PA) is

$$\frac{24}{75} = 0.32,$$

or 32.0%. In the synthetic phonics group (SP), the proportion is

$$\frac{78}{113} \simeq 0.690,$$

or about 69.0%.

- (b) The proportions in the three groups are: AP: 35%; AP + PA: 32%; and SP: 69%. The proportions of children whose reading age exceeds their chronological age remained steady, or perhaps even dropped, in the two groups receiving analytic phonics, whereas it increased dramatically (to 69% from 35% at the baseline test) in the synthetic phonics group.
- (c) In order to interpret these results, it is necessary to find out whether the differences between the groups at the first follow-up test could be due to chance fluctuations. You might also want to know whether the impact of the synthetic phonics method is sustained over time (and indeed how the children performed on other tests).

Solution to Activity 5

- (a) This is not a contingency table. The column variable, chronological age at test, is not categorical. Note that it therefore makes no sense to consider whether the categories are mutually exclusive, as there are no categories. There are two columns, but these represent two different measurements on the same variable (age), rather than two categories of the same variable.
- (b) This is not a contingency table. Although both variables are categorical and the entries are counts, the categories across the columns are not mutually exclusive: a child can have scored well both at the baseline test and at the first follow-up test.

Solution to Activity 6

- (a) The row variable has 3 categories and the column variable has 4 categories. Thus, $r = 3$ and $c = 4$, and so this is a 3×4 contingency table.
- (b) Table 4 is 3×2 . Swapping the rows and columns around would turn it into a 2×3 contingency table.

Solution to Activity 7

(a) This is the joint probability of events R_2 and SP .

$$\begin{aligned} P(R_2 \text{ and } SP) &= \frac{\text{number of children with event } R_2 \text{ and event } SP}{\text{total number of children}} \\ &= \frac{90}{268} \simeq 0.336. \end{aligned}$$

(b) This is the conditional probability of R_2 , given SP .

$$\begin{aligned} P(R_2|SP) &= \frac{\text{number of children experiencing } R_2 \text{ and } SP}{\text{number of children experiencing } SP} \\ &= \frac{90}{105} \simeq 0.857. \end{aligned}$$

(c) This is the conditional probability of $AP+$, given R_2 .

$$\begin{aligned} P(AP+|R_2) &= \frac{\text{number of children experiencing } AP+ \text{ and } R_2}{\text{number of children experiencing } R_2} \\ &= \frac{55}{223} \simeq 0.247. \end{aligned}$$

Solution to Activity 8

(a) $P(R_2|SP)$

(b) $P(R_2 \text{ and } SP)$

(c) $P(SP|R_2)$

(d) $P(R_2 \text{ and } SP)$

Solution to Activity 9

For sixth form students, let 'Phys' denote the event 'studies physics', and 'Chem' the event 'studies chemistry'. Since 30% of sixth form students study physics, we have

$$P(\text{Phys}) = 0.30.$$

Also, among students who study physics, 80% also study chemistry, so

$$P(\text{Chem}|\text{Phys}) = 0.80.$$

Hence the proportion of students who study both physics and chemistry is

$$\begin{aligned} P(\text{Phys and Chem}) &= P(\text{Phys}) \times P(\text{Chem}|\text{Phys}) \\ &= 0.30 \times 0.80 \\ &= 0.24. \end{aligned}$$

Hence the probability that a student picked at random studies both physics and chemistry is 0.24.

Solution to Activity 10

(a) By direct calculation, we have

$$P(S_0|AP+) = \frac{35}{78} \simeq 0.449,$$

and

$$P(S_0|SP) = \frac{47}{117} \simeq 0.402.$$

(b) We have, retaining more decimal places for intermediate results,

$$P(S_0 \text{ and } AP+) = \frac{35}{303} \simeq 0.1155$$

and

$$P(AP+) = \frac{78}{303} \simeq 0.2574.$$

So

$$P(S_0|AP+) \simeq \frac{0.1155}{0.2574} \simeq 0.449,$$

as obtained in part (a). Similarly,

$$P(S_0 \text{ and } SP) = \frac{47}{303} \simeq 0.1551$$

and

$$P(SP) = \frac{117}{303} \simeq 0.3861.$$

So

$$P(S_0|SP) \simeq \frac{0.1551}{0.3861} \simeq 0.402,$$

which is also equal to the corresponding value in part (a).

(c) The percentages spelling better than expected are therefore about 47.2% in the AP group, 44.9% in the $AP+$ group, and 40.2% in the SP group. The proportions are therefore broadly similar across groups, though a little lower in the SP group.

Solution to Activity 11

(a) Let ‘Phys’ denote the event ‘studies physics’, and ‘Chem’ the event ‘studies chemistry’. Then

$$\begin{aligned} P(\text{Phys or Chem}) &= P(\text{Phys}) + P(\text{Chem}) - P(\text{Phys and Chem}) \\ &= 0.30 + 0.32 - 0.24 \\ &= 0.38. \end{aligned}$$

So 38% of sixth form students study physics or chemistry at this school.

(b) Let R_0 denote the event ‘reading age at baseline higher than chronological age’. Then

$$P(R_0) = \frac{110}{291} \simeq 0.378, \quad P(R_1) = \frac{138}{291} \simeq 0.474,$$

and

$$P(R_0 \text{ and } R_1) = \frac{60}{291} \simeq 0.206.$$

So, using the general rule for adding probabilities,

$$\begin{aligned} P(R_0 \text{ or } R_1) &= P(R_0) + P(R_1) - P(R_0 \text{ and } R_1) \\ &\simeq 0.378 + 0.474 - 0.206 \\ &= 0.646. \end{aligned}$$

(c) Rearranging the general rule for adding probabilities,

$$P(A \text{ and } B) = P(A) + P(B) - P(A \text{ or } B).$$

Thus,

$$\begin{aligned} P(R_1 \text{ and } R_2) &= P(R_1) + P(R_2) - P(R_1 \text{ or } R_2) \\ &= 0.483 + 0.840 - 0.848 \\ &= 0.475. \end{aligned}$$

So $263 \times 0.475 \simeq 125$ children did well on both tests.

(If unrounded values for $P(R_1)$, $P(R_2)$ and $P(R_1 \text{ or } R_2)$ are used instead of values correct to three decimal places as above, the product $263 \times P(R_1 \text{ and } R_2)$ is exactly equal to 125 and no ‘ $\simeq$ ’ is needed.)

Solution to Activity 12

(a) The probabilities are

$$P(S_2|AP) = \frac{85}{95} \simeq 0.895,$$

$$P(S_2|AP+) = \frac{59}{66} \simeq 0.894,$$

and

$$P(S_2|SP) = \frac{94}{104} \simeq 0.904.$$

(b) Using the marginal totals,

$$P(S_2) = \frac{238}{265} \simeq 0.898.$$

(c) The three conditional probabilities are very close to each other, and to the marginal probability $P(S_2)$, though not identical. This suggests that the original teaching group has little influence on the outcome of the second spelling test. It appears that the children in the two analytic phonics groups ‘caught up’ with those in the synthetic phonics group, after the synthetic phonics programme was extended to all children in the study.

Solution to Activity 13

- (a) We have $P(E \text{ and } M) = 0.48$, directly from Table 17.

The proportion studying English equals the proportion studying English and maths, plus the proportion studying English but not maths. So

$$P(E) = 0.48 + 0.32 = 0.80.$$

Similarly the proportion studying maths equals the proportion studying English and maths, plus the proportion studying maths but not English. So

$$P(M) = 0.48 + 0.12 = 0.60.$$

- (b) We have

$$P(E|M) = \frac{P(E \text{ and } M)}{P(M)} = \frac{0.48}{0.60} = 0.80,$$

and

$$P(M|E) = \frac{P(M \text{ and } E)}{P(E)} = \frac{0.48}{0.80} = 0.60.$$

- (c) Notice that $P(E|M) = P(E) = 0.80$, and $P(M|E) = P(M) = 0.60$. Thus, the events E and M are statistically independent.

Solution to Activity 14

The hypotheses are:

H_0 : Teaching method for children starting at primary school and reading ability at the baseline test are independent.

H_1 : Teaching method for children starting at primary school and reading ability at the baseline test are not independent.

Solution to Activity 15

- (a) For the analytic phonics + PA group, the Expected values are:

$$\text{Not higher: } 75 \times \frac{154}{292} \simeq 39.5548, \quad \text{Higher: } 75 \times \frac{138}{292} \simeq 35.4452.$$

- (b) For the synthetic phonics group, the Expected values are:

$$\text{Not higher: } 113 \times \frac{154}{292} \simeq 59.5959, \quad \text{Higher: } 113 \times \frac{138}{292} \simeq 53.4041.$$

Solution to Activity 16

Start by copying the marginal totals from the Observed table to the Expected table. Then complete the Expected table by applying the equation in the highlighted box. We obtain, for example, the following Expected value for the ‘Not higher’ cell in the ‘analytic phonics’ group:

$$E = \frac{108 \times 188}{303} \simeq 67.0099.$$

The other Expected values are calculated in a similar manner.

	Not higher	Higher	Total
Analytic phonics	67.0099	40.9901	108
Analytic phonics + PA	48.3960	29.6040	78
Synthetic phonics	72.5941	44.4059	117
Total	188	115	303

The Expected values do indeed add up to the marginal totals from the Observed table.

Solution to Activity 17

- (a) The χ^2 contribution for the cell corresponding to ‘Not higher’ in the analytic phonics group is

$$\chi^2 \text{ contribution} = \frac{(69 - 67.0099)^2}{67.0099} \simeq 0.0591.$$

The other contributions are calculated in similar fashion, and are shown in the following table.

	Not higher	Higher
Analytic phonics	0.0591	0.0966
Analytic phonics + PA	0.6016	0.9835
Synthetic phonics	0.1598	0.2612

- (b) The χ^2 test statistic is:

$$\begin{aligned} \chi^2 &= 0.0591 + 0.0966 + 0.6016 + 0.9835 + 0.1598 + 0.2612 \\ &= 2.1618 \\ &\simeq 2.162. \end{aligned}$$

Solution to Activity 18

We have $\chi^2 = 2.162 < 5.991$. Since $\chi^2 < CV_5$, do not reject the null hypothesis. Thus, there is little evidence against the null hypothesis that reading ability in the baseline test is independent of the teaching group to which the children in the study were allocated.

Solution to Activity 19

- (a) Degrees of freedom = 1; $CV5 = 3.841$, $CV1 = 6.635$.
 (b) Degrees of freedom = 6; $CV5 = 12.592$, $CV1 = 16.812$.
 (c) Degrees of freedom = 12; $CV5 = 21.026$, $CV1 = 26.217$.
 (d) Degrees of freedom = 3; $CV5 = 7.815$, $CV1 = 11.345$.

Solution to Activity 20

- (a) The hypotheses are as follows:

H_0 : Teaching method originally allocated and reading ability at the second follow-up test are independent.

H_1 : Teaching method originally allocated and reading ability at the second follow-up test are not independent.

- (b) Copy marginal totals from the Observed table to the Expected table. Then the Expected value for the first cell is

$$\frac{\text{row total} \times \text{column total}}{\text{overall total}} = \frac{97 \times 45}{268} \simeq 16.2873.$$

The other values are obtained in the same manner, leading to the following Expected table:

	Not higher	Higher	Total
Analytic phonics	16.2873	80.7127	97
Analytic phonics + PA	11.0821	54.9179	66
Synthetic phonics	17.6306	87.3694	105
Total	45	223	268

As a check on your calculations, the expected values within the table must add to the marginal totals (apart from unimportant rounding errors).

- (c) The Residual table is found by subtracting the terms of the Expected table from those of the Observed table. For the first cell, we have

$$19 - 16.2873 = 2.7127.$$

The Residual table is:

	Not higher	Higher
Analytic phonics	2.7127	-2.7127
Analytic phonics + PA	-0.0821	0.0821
Synthetic phonics	-2.6306	2.6306

The χ^2 contribution of the first cell is

$$\frac{\text{Residual}^2}{\text{Expected}} = \frac{2.7127^2}{16.2873} \simeq 0.4518.$$

The complete table of χ^2 contributions is:

	Not higher	Higher
Analytic phonics	0.4518	0.0912
Analytic phonics + PA	0.0006	0.0001
Synthetic phonics	0.3925	0.0792

- (d) The χ^2 test statistic is the sum of the six χ^2 contributions:

$$\chi^2 = 0.4518 + 0.0912 + 0.0006 + 0.0001 + 0.3925 + 0.0792 \\ \simeq 1.015.$$

The Observed table is a 3×2 table, so its degrees of freedom are $(3 - 1) \times (2 - 1) = 2$. Hence $CV_5 = 5.991$ and $CV_1 = 9.210$.

- (e) Since $1.015 < 5.991$, we do not reject the null hypothesis that the teaching method originally allocated and reading ability at the second follow-up test are independent. In other words, there is little evidence that children's reading ability at this stage differs between the three teaching groups.

Solution to Activity 21

- (a) The hypotheses are as follows:

H_0 : Teaching method and ability in spelling/reading at the first follow-up test are independent.

H_1 : Teaching method and ability in spelling/reading at the first follow-up test are not independent.

- (b) Copy marginal totals from the Observed table to the Expected table. The Expected value for the first cell is

$$\frac{\text{row total} \times \text{column total}}{\text{overall total}} = \frac{179 \times 136}{292} \simeq 83.3699.$$

The other values are obtained in the same manner, leading to the following Expected table (with Observed marginal totals):

	Low/Low	Low/High	High/Low	High/High	Total
Any AP	83.3699	17.7774	11.0342	66.8185	179
Synthetic phonics	52.6301	11.2226	6.9658	42.1815	113
Total	136	29	18	109	292

As a check on your calculations, the expected values within the table must add to the marginal totals (apart from unimportant rounding errors).

All Expected values are indeed ≥ 5 .

- (c) The Residual table is found by subtracting the terms of the Expected table from those of the Observed table. For the first cell, we have

$$107 - 83.3699 = 23.6301.$$

The Residual table is:

	Low/Low	Low/High	High/Low	High/High
Any AP	23.6301	4.2226	0.9658	-28.8185
Synthetic phonics	-23.6301	-4.2226	-0.9658	28.8185

The χ^2 contribution of the first cell is

$$\frac{\text{Residual}^2}{\text{Expected}} = \frac{23.6301^2}{83.3699} \simeq 6.6976.$$

The complete table of χ^2 contributions is:

	Low/Low	Low/High	High/Low	High/High
Any AP	6.6976	1.0030	0.0845	12.4293
Synthetic phonics	10.6095	1.5888	0.1339	19.6889

- (d) The χ^2 test statistic is the sum of the eight χ^2 contributions:

$$\begin{aligned}\chi^2 &= 6.6976 + 1.0030 + 0.0845 + 12.4293 + 10.6095 \\ &\quad + 1.5888 + 0.1339 + 19.6889 \\ &\simeq 52.236.\end{aligned}$$

The Observed table is a 2×4 table, so its degrees of freedom are $(2 - 1) \times (4 - 1) = 3$. Hence $CV_5 = 7.815$ and $CV_1 = 11.345$.

- (e) Since $52.236 \geq 11.345$, we reject at the 1% significance level the null hypothesis that teaching method and spelling/reading ability at the first follow-up test are independent. There is strong evidence that children's spelling/reading ability differs between the two teaching groups.
- (f) Examination of the χ^2 contributions shows that the two biggest values are for the High/High ability category. The residuals corresponding to these cells are negative for 'any analytic phonics' and positive for synthetic phonics. Thus, the major departure from independence is due to children in the synthetic phonics group performing better in both spelling and reading than children in the 'any analytic phonics' group.

Solution to Activity 22

- (a) The Expected table (with Observed marginal totals) is:

	Not higher	Higher	Total
Analytic phonics	85.2290	117.7710	203
Synthetic phonics	79.7710	110.2290	190
Total	165	228	393

The Residual table is:

	Not higher	Higher
Analytic phonics	29.7710	-29.7710
Synthetic phonics	-29.7710	29.7710

The χ^2 contributions table is:

	Not higher	Higher
Analytic phonics	10.3992	7.5257
Synthetic phonics	11.1107	8.0406

The χ^2 test statistic is:

$$\chi^2 = 10.3992 + 7.5257 + 11.1107 + 8.0406 \simeq 37.076.$$

The Observed table has $(2 - 1) \times (2 - 1) = 1$ degree of freedom, so $CV_5 = 3.841$ and $CV_1 = 6.635$. Since $37.076 \geq 6.635$, we reject the null hypothesis at the 1% significance level. Thus, there is strong evidence that reading ability varies with teaching method.

- (b) Examination of the Observed table (you could also look at the χ^2 contributions and residuals, but for a 2×2 table it's easier to just look at the data) shows that the synthetic phonics group have done better than the analytic phonics group. If the impact of synthetic phonics teaching was short-lived, we would expect to see no difference between the groups. Thus, on the face of it, it seems that the data do not support this hypothesis. Thus, synthetic phonics teaching appears to confer a lasting advantage.

Solution to Activity 23

- (a) The Expected table (with Observed marginals) is:

	Not higher	Higher	Total
Analytic phonics	41.5721	67.4279	109
Synthetic phonics	40.4279	65.5721	106
Total	82	133	215

The Residual table is:

	Not higher	Higher
Analytic phonics	20.4279	-20.4279
Synthetic phonics	-20.4279	20.4279

The χ^2 contributions table is:

	Not higher	Higher
Analytic phonics	10.0380	6.1888
Synthetic phonics	10.3221	6.3640

The χ^2 test statistic is 32.913. This is greater than 6.635, the value of CV1 for one degree of freedom. Hence we reject the null hypothesis that boys' reading ability is independent of teaching group.

Examination of the Observed table shows that boys in the synthetic phonics group perform better than boys in the analytic phonics group.

- (b) The Expected table (with Observed marginals) is:

	Not higher	Higher	Total
Analytic phonics	43.8315	50.1685	94
Synthetic phonics	39.1685	44.8315	84
Total	83	95	178

The Residual table is:

	Not higher	Higher
Analytic phonics	9.1685	-9.1685
Synthetic phonics	-9.1685	9.1685

The χ^2 contributions table is:

	Not higher	Higher
Analytic phonics	1.9178	1.6756
Synthetic phonics	2.1461	1.8751

The χ^2 test statistic is 7.615. This is greater than 6.635, the value of CV1 for one degree of freedom. Hence we reject the null hypothesis that girls' reading ability is independent of teaching group.

Examination of the Observed table shows that girls in the synthetic phonics group perform better than girls in the analytic phonics group.

- (c) Reading ability is better in the synthetic phonics group than in the analytic phonics group for both boys and girls. Thus, gender is not a confounder for the association between reading ability and teaching method.

Solution to Activity 24

- (a) The null hypothesis of independence is false, but was not rejected: a type 2 error was made.
- (b) The null hypothesis of independence is false, and was rejected: no error was made.
- (c) The null hypothesis of independence is true, but was rejected: a type 1 error was made.
- (d) The null hypothesis is true, and was not rejected: no error was made.

Solutions to exercises

Solution to Exercise 1

One way to look at this would be to do separate analyses for boys and girls within each teaching method group, to see if there are differences between boys and girls, and whether those differences vary between teaching methods.

Solution to Exercise 2

- (a) The overall proportion of girls is

$$\frac{146}{304} \simeq 0.480.$$

Hence the percentage of girls is 48.0% (to one decimal place).

- (b) The proportion of boys in the analytic phonics group is

$$\frac{58}{109} \simeq 0.532.$$

In the analytic phonics + PA group it is

$$\frac{39}{78} = 0.5,$$

while in the synthetic phonics group it is

$$\frac{61}{117} \simeq 0.521.$$

- (c) The proportions of boys in the three groups are similar. (You would not expect them to be exactly equal, owing to random fluctuations.)

Solution to Exercise 3

- (a) Table 10 is a contingency table. The variables are categorical: the row variable has two mutually exclusive categories, ‘analytic phonics with or without PA’ and ‘synthetic phonics’, and gender also has two mutually exclusive categories (for our purposes, anyway). Finally, the entries are counts.

Table 11 is not a contingency table: as made clear in its caption, it represents percentages, so the entries are not counts. The percentages have been rounded to the nearest whole number so they might *look* like counts, but they are not (and they add up to 100% for each row).

- (b) Each of the variables in Table 10 has two categories, so the table is a 2×2 table. Table 11 is not a contingency table, but if you thought it was and calculated its dimension as if it were a contingency table, you’d most likely have observed that it has 3 rows and 2 columns, and so classified it as being a 3×2 table.

Solution to Exercise 4

(a) $P(G|AP) = \frac{51}{109} \simeq 0.468.$

(b) $P(AP+) = \frac{78}{304} \simeq 0.257.$

(c) $P(SP|B) = \frac{61}{158} \simeq 0.386.$

(d) $P(B \text{ and } AP+) = \frac{39}{304} \simeq 0.128.$

(e) This probability is

$$P(G|SP) = \frac{56}{117} \simeq 0.479.$$

(f) This probability is

$$P(AP|B) = \frac{58}{158} \simeq 0.367.$$

Solution to Exercise 5

(a) We have

$$P(R_2|S_2) = \frac{P(R_2 \text{ and } S_2)}{P(S_2)}.$$

Now, retaining more decimal places for intermediate results,

$$P(S_2) = \frac{237}{264} \simeq 0.8977,$$

and

$$P(R_2 \text{ and } S_2) = \frac{214}{264} \simeq 0.8106.$$

So

$$P(R_2|S_2) \simeq \frac{0.8106}{0.8977} \simeq 0.903.$$

(b) This probability is $P(S_2|R_2)$. Thus, similarly to part (a),

$$P(S_2|R_2) = \frac{P(S_2 \text{ and } R_2)}{P(R_2)}.$$

Now

$$P(R_2) = \frac{220}{264} \simeq 0.8333,$$

and (from the solution to part (a))

$$P(S_2 \text{ and } R_2) \simeq 0.8106.$$

So

$$P(S_2|R_2) \simeq \frac{0.8106}{0.8333} \simeq 0.973.$$

- (c) Applying the general addition rule,

$$P(R_2 \text{ or } S_2) = P(R_2) + P(S_2) - P(R_2 \text{ and } S_2).$$

From parts (a) and (b), $P(R_2) \simeq 0.8333$, $P(S_2) \simeq 0.8977$, and $P(R_2 \text{ and } S_2) \simeq 0.8106$. So

$$P(R_2 \text{ or } S_2) \simeq 0.8333 + 0.8977 - 0.8106 \simeq 0.920.$$

Solution to Exercise 6

- (a) The row and column variables are categorical. The variables each have mutually exclusive categories, since a child must be either ‘Low’ or ‘High’ on each test. Finally, the entries are counts. Hence Table 19 is a contingency table.
- (b) We have

$$P(\text{Reading High}|\text{Spelling Low}) = \frac{6}{27} \simeq 0.222,$$

and

$$P(\text{Reading High}) = \frac{220}{264} \simeq 0.833.$$

- (c) The two probabilities are very different. This suggests that spelling and reading ability are not independent. However, some difference is to be expected owing to random fluctuations. We need to know whether the observed difference could plausibly have arisen by chance.

Solution to Exercise 7

- (a) The hypotheses are as follows:

H_0 : Teaching method and spelling ability at the first follow-up test are independent.

H_1 : Teaching method and spelling ability at the first follow-up test are not independent.

- (b) Copy marginal totals from the Observed table to the Expected table. The Expected value for the first cell is

$$\frac{\text{row total} \times \text{column total}}{\text{overall total}} = \frac{104 \times 165}{292} \simeq 58.7671.$$

The other values are obtained in the same manner, leading to the following Expected table:

	Not higher	Higher	Total
Analytic phonics	58.7671	45.2329	104
Analytic phonics + PA	42.3801	32.6199	75
Synthetic phonics	63.8527	49.1473	113
Total	165	127	292

(To check your calculations, add up the expected values in each column and in each row. These should give the marginal totals of the table, apart from unimportant rounding errors.)

- (c) The Residual table is found by subtracting the terms of the Expected table from those of the Observed table. For the first cell, we have

$$79 - 58.7671 = 20.2329.$$

The Residual table is:

	Not higher	Higher
Analytic phonics	20.2329	-20.2329
Analytic phonics + PA	7.6199	-7.6199
Synthetic phonics	-27.8527	27.8527

The χ^2 contribution of the first cell is

$$\frac{\text{Residual}^2}{\text{Expected}} = \frac{20.2329^2}{58.7671} \simeq 6.9660.$$

The complete table of χ^2 contributions is:

	Not higher	Higher
Analytic phonics	6.9660	9.0503
Analytic phonics + PA	1.3701	1.7800
Synthetic phonics	12.1494	15.7846

- (d) The χ^2 test statistic is the sum of the six χ^2 contributions:

$$\begin{aligned}\chi^2 &= 6.9660 + 9.0503 + 1.3701 + 1.7800 + 12.1494 + 15.7846 \\ &\simeq 47.100.\end{aligned}$$

The Observed table is a 3×2 table, so its degrees of freedom are $(3 - 1) \times (2 - 1) = 2$. Hence $CV_5 = 5.991$ and $CV_1 = 9.210$.

- (e) Since $47.100 \geq 9.210$, we reject at the 1% significance level the null hypothesis that teaching method and spelling ability at the first follow-up test are independent. There is strong evidence that children's spelling ability at this stage differs between the three teaching groups.
- (f) Examination of the table of χ^2 contributions shows that the largest values are for the synthetic phonics group. The corresponding residuals are negative for 'Not higher' and positive for 'Higher'. Thus, the major departure from independence comes from children in the synthetic phonics group performing better in the spelling test than would be expected under the assumption of independence.

Solution to Exercise 8

- (a) The hypotheses are as follows:

H_0 : Teaching method originally allocated and spelling ability at the second follow-up test are independent.

H_1 : Teaching method originally allocated and spelling ability at the second follow-up test are not independent.

- (b) Copy marginal totals from the Observed table to the Expected table. The Expected value for the first cell is

$$\frac{\text{row total} \times \text{column total}}{\text{overall total}} = \frac{95 \times 27}{265} \simeq 9.6792.$$

The other values are obtained in the same manner, leading to the following Expected table:

	Not higher	Higher	Total
Analytic phonics	9.6792	85.3208	95
Analytic phonics + PA	6.7245	59.2755	66
Synthetic phonics	10.5962	93.4038	104
Total	27	238	265

As a check on your calculations, the expected values within the table must add to the marginal totals (ignoring unimportant rounding errors).

- (c) The Residual table is found by subtracting the terms of the Expected table from those of the Observed table. For the first cell, we have

$$10 - 9.6792 = 0.3208.$$

The Residual table is:

	Not higher	Higher
Analytic phonics	0.3208	-0.3208
Analytic phonics + PA	0.2755	-0.2755
Synthetic phonics	-0.5962	0.5962

The χ^2 contribution of the first cell is

$$\frac{\text{Residual}^2}{\text{Expected}} = \frac{0.3208^2}{9.6792} \simeq 0.0106.$$

The complete table of χ^2 contributions is:

	Not higher	Higher
Analytic phonics	0.0106	0.0012
Analytic phonics + PA	0.0113	0.0013
Synthetic phonics	0.0335	0.0038

- (d) The χ^2 test statistic is the sum of the six χ^2 contributions:

$$\chi^2 = 0.0106 + 0.0012 + 0.0113 + 0.0013 + 0.0335 + 0.0038 \simeq 0.062.$$

The Observed table is a 3×2 table, so its degrees of freedom are $(3 - 1) \times (2 - 1) = 2$. Hence $CV_5 = 5.991$ and $CV_1 = 9.210$.

- (e) Since $0.062 < 5.991$, we do not reject the null hypothesis that teaching method and spelling ability at the second follow-up test are independent. There is little evidence that children's spelling ability at

this stage is related to the groups to which they were originally allocated.

Solution to Exercise 9

(a) The hypotheses are as follows:

H_0 : Teaching method and ability in spelling/reading at the first follow-up test are independent.

H_1 : Teaching method and ability in spelling/reading at the first follow-up test are not independent.

(b) Copy marginal totals from the Observed table to the Expected table. The Expected value for the first cell is

$$\frac{\text{row total} \times \text{column total}}{\text{overall total}} = \frac{104 \times 136}{292} \simeq 48.4384.$$

The other values are obtained in the same manner, leading to the following Expected table:

	Low/Low	L/H or H/L	High/High	Total
Analytic phonics	48.4384	16.7397	38.8219	104
Analytic phonics + PA	34.9315	12.0719	27.9966	75
Synthetic phonics	52.6301	18.1884	42.1815	113
Total	136	47	109	292

As a check on your calculations, the expected values within the table must add to the marginal totals (ignoring unimportant rounding errors).

(c) The Residual table is found by subtracting the terms of the Expected table from those of the Observed table. For the first cell, we have

$$65 - 48.4384 = 16.5616.$$

The Residual table is:

	Low/Low	L/H or H/L	High/High
Analytic phonics	16.5616	0.2603	-16.8219
Analytic phonics + PA	7.0685	4.9281	-11.9966
Synthetic phonics	-23.6301	-5.1884	28.8185

The χ^2 contribution of the first cell is

$$\frac{\text{Residual}^2}{\text{Expected}} = \frac{16.5616^2}{48.4384} \simeq 5.6626.$$

The complete table of χ^2 contributions is:

	Low/Low	L/H or H/L	High/High
Analytic phonics	5.6626	0.0040	7.2891
Analytic phonics + PA	1.4303	2.0118	5.1406
Synthetic phonics	10.6095	1.4800	19.6889

- (d) The χ^2 test statistic is the sum of the nine χ^2 contributions:

$$\begin{aligned}\chi^2 &= 5.6626 + 0.0040 + 7.2891 + 1.4303 + 2.0118 + 5.1406 \\ &\quad + 10.6095 + 1.4800 + 19.6889 \\ &\simeq 53.317.\end{aligned}$$

The Observed table is a 3×3 table, so its degrees of freedom are $(3 - 1) \times (3 - 1) = 4$. Hence $CV_5 = 9.488$ and $CV_1 = 13.277$.

- (e) Since $53.317 \geq 13.277$, we reject at the 1% significance level the null hypothesis that teaching method and spelling/reading ability at the first follow-up test are independent. There is strong evidence that children's spelling/reading ability differs between the three teaching groups.
- (f) Examination of the χ^2 contributions shows that the two biggest values are for the High/High and Low/Low ability categories within the synthetic phonics group. The residuals corresponding to these cells are negative for Low/Low and positive for High/High. Thus, the major departure from independence is due to children in the synthetic phonics group performing better in both spelling and reading than children in the other teaching method groups.

Solution to Exercise 10

- (a) The contingency table with marginal totals is:

	Height above ground		
	3 feet	35 feet	Total
Males	173	125	298
Females	150	73	223
Total	323	198	521

The hypotheses are as follows:

H_0 : The gender of sandflies caught in light-traps is independent of the height above the ground of the traps.

H_1 : The gender of sandflies caught in light-traps is not independent of the height above the ground of the traps.

- (b) The Expected table (with Observed marginal totals) is:

	3 feet	35 feet	Total
Males	184.7486	113.2514	298
Females	138.2514	84.7486	223
Total	323	198	521

Note that all Expected values are greater than or equal to 5, so the test can proceed.

The Residual table is:

	3 feet	35 feet
Males	-11.7486	11.7486
Females	11.7486	-11.7486

The χ^2 contributions table is:

	3 feet	35 feet
Males	0.7471	1.2188
Females	0.9984	1.6287

The χ^2 test statistic is 4.593. This is greater than 3.841, the value of CV5 for one degree of freedom, but not greater than CV1, which is 6.635. Hence we reject the null hypothesis at the 5% significance level, but not at the 1% significance level.

We conclude that there is some evidence that the gender of sandflies caught in light-traps and the height of traps above the ground are not independent.

None of the terms in the χ^2 contributions table are large. We note that the number of males caught in traps 35 feet above ground level and the number of females caught in traps 3 feet above ground level are higher than the corresponding Expected values. In fact 42% ($125/298 \times 100\%$) of males were caught 35 feet above ground level, compared to only 33% of females. This suggests that male sandflies are more likely to fly higher than females.

- (c) Since the null hypothesis has been rejected, a type 1 error might have been committed.

Solution to Exercise 11

- (a) The null and alternative hypotheses are as follows.

H_0 : The blood group of an individual in Iraq is independent of which population group he or she belongs to.

H_1 : The blood group of an individual in Iraq is not independent of which population group he or she belongs to.

- (b) The Expected table is as follows, with Observed marginal totals.

	O	A	B	AB	Total
Kurd	533.5019	463.9697	316.0556	186.4728	1500
Arab	175.3443	152.4914	103.8769	61.2874	493
Other	135.1538	117.5390	80.0674	47.2398	380
Total	844	734	500	295	2373

All Expected values are greater than or equal to 5, so the test can proceed.

The Residual table is as follows:

	O	A	B	AB
Kurd	-2.5019	-13.9697	-23.0556	39.5272
Arab	-1.3443	-2.4914	29.1231	-25.2874
Other	3.8462	16.4610	-6.0674	-14.2398

The χ^2 contributions tables is:

	O	A	B	AB
Kurd	0.0117	0.4206	1.6819	8.3787
Arab	0.0103	0.0407	8.1650	10.4337
Other	0.1095	2.3053	0.4598	4.2924

The χ^2 test statistic is $\chi^2 = 36.310$.

The degrees of freedom are $(3 - 1) \times (4 - 1) = 6$, so $CV_5 = 12.592$ and $CV_1 = 16.812$. Since $36.310 \geq 16.812$, we reject H_0 at the 1% significance level. There is strong evidence that blood group is not independent of population group for people in Iraq.

- (c) We might have committed a type 1 error, as we rejected the null hypothesis.

The highest χ^2 terms are 8.3787, 8.1650 and 10.4337, which come from the Kurd and Arab populations. Of the Kurds, 15% are of group AB compared to only 7% of Arabs. Arabs have a much higher percentage (27%) of group B than the other population groups.

Acknowledgements

Grateful acknowledgement is made to the following sources:

Cover image: Minxlj/www.flickr.com/photos/minxlj/422472167/. This file is licensed under the Creative Commons Attribution-Non commercial-No Derivatives Licence <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Figure 1 © Penguin Group

Figure 2 TUBS. http://en.wikipedia.org/wiki/File:Clackmannanshire_in_Scotland.svg. This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

Exercises on Section 1, cartoon: Randy Glasbergen / www.glasbergen.com

Subsection 2.1 photo ('Learning how to read'), Monkey Business Images / Dreamstime

Subsection 2.2 photo ('Children reading'), Cliff Parnell / www.istockphoto.com

Subsection 3.1 cartoon (reading at a 3rd grade level), www.cartoonstock.com

Subsection 3.2 photo ('In the classroom'), daaronj / www.istockphoto.com

Subsection 5.2 cartoon (20 years experience), www.causeweb.org

Subsection 5.2 cartoon (question the data), www.causeweb.org

Exercises on Section 5, photo of sandfly: Commander Applebery / Flickr.com. This file is licensed under the Creative Commons Attribution-Share Alike Licence <http://creativecommons.org/licenses/by-sa/3.0/>

Every effort has been made to contact copyright holders. If any have been inadvertently overlooked the publishers will be pleased to make the necessary arrangements at the first opportunity.